

Access DB# 84028**SEARCH REQUEST FORM**

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval  
Reference Librarian  
Biotechnology & Chemical Library  
CM1 1E07 - 703-308-4498  
jan.delaval@uspto.gov

\*\*\*\*\*  
**STAFF USE ONLY****Type of Search****Vendors and cost where applicable**

Searcher: <u>Jan</u>	NA Sequence (#) _____	STN _____
Searcher Phone #: <u>1448</u>	AA Sequence (#) <input checked="" type="checkbox"/> _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>11/12/03</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>11/12/03</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <input checked="" type="checkbox"/> _____
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____





```
QY 241 DEVVLVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
Db 241 DEVVLVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
QY 301 FGSELSPTATPTSLKLGVAIDICVVLASSSEATTSQELRLRVQGEKHQMLEISLSP 360
Db 301 FGSELSPTATPTSLKLGVAIDICVVLASSSEATTSQELRLRVQGEKHQMLEISLSP 360
QY 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGTLKSGKELPADLGESGDLIEVWG 412
Db 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGTLKSGKELPADLGESGDLIEVWG 412

RESULT 2
US-09-175-254-3
; Sequence 3, Application US/09175254
; Patent No. US20020028482A1
; GENERAL INFORMATION:
; APPLICANT: Hong Zhou
; APPLICANT: Jiuqiao Zhao
; APPLICANT: Derong Liu
; TITLE OF INVENTION: NIP45 HUMAN HOMOLOG
; FILE REFERENCE: PHM 70281
; CURRENT APPLICATION NUMBER: US/09/175,254
; EARLIER FILING DATE: 1998-10-20
; EARLIER APPLICATION NUMBER: GB 9722388.7
; EARLIER FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-175-254-3
```

```
Query Match 66.7%; Score 1399.5; DB 10; Length 419;
Best Local Similarity 70.1%; Pred. No. 6.8e-93;
Matches 295; Conservative 30; Mismatches 85; Indels 11; Gaps 4;

QY 1 MAEPLRGRGPRSRG---GRCARRARGRCPRARQSPARLIPDTVLVDLVSDSEVLE 57
Db 1 MAEPLVGRGWRGSGGAGRGGRGWGGRPRRAQRSPSRGTLDVSVDLVTDSDEILLE 60
QY 58 V-----ADPVEVPVARLPAPAKEQSDSDSEGAAGPAGAPRTLVRRRRRLDPGEA 111
Db 61 VATARGAADEVVEPEPPPGFVASRDNNSGDEGDRRPPAGPREPVRRRRLVLDPGEA 120
QY 112 PVVPVYSGKVQSSNLIPDSSLLKLCPSPEDEADLTNSGSSPSEDDALPSPGWRKLL 171
Db 121 PLVPVYSGKVYSSLRLLIPDLSLLKLYPPGDEBEAEALDSSGLYHEGSPSP-GSPWKTL 179
QY 172 RKCEKEEKMEPEPDQDISPLPOPSSRNKSRKHTALQKLRVNRKLDLRSLCSPKQH 231
Db 180 RTK-DKEEKKKTEFLDNDGFLPSPSPRTKSRTHTRALKLSEVNRKLDLRSLCSPKPP 238
QY 232 QSPALQSTDDDEVVLVEGVPVLPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 291
Db 239 QGQEQOQOQDEVLVEGTPETLPTLRLPFLKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 298
QY 292 VSPNRILLFGESLSPTATPTSLKLGVAIDICVVLASSSEATTSQELRLRVQGEKH 351
Db 299 VSPSRILLFGETLSPTATPTSLKLGVAIDICVVLASSSEATTSQELRLRVQGEKH 358
QY 352 QMLEISLSPSLKVLMSHYEAMGLSGHKLSPFFDGTGTLKSGKELPADLGESGDLIEVW 411
Db 359 QTLVLSRSLSPSLKVLMSHYEAMGLSGHKLSPFFDGTGTLKSGKELPADLGESGDLIEVW 418
QY 412 G 412
Db 419 G 419

RESULT 3
```

```
US-09-925-299-862
; Sequence 862, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862
```

```
Query Match 5.9%; Score 124.5; DB 10; Length 518;
Best Local Similarity 21.4%; Pred. No. 0.13;
Matches 101; Conservative 48; Mismatches 150; Indels 174; Gaps 22;

QY 8 RGRSRGGGRRARGRCPRARQSPARLIPDTVLVDLV-----SDSD 52
Db 97 RAPR-RGPLGGRKKK----KAPSASDSKADSDGAKPEPVMARSAASSSSSSSSSDSD 151
QY 53 BEVLE-----VADPEVEVPVARLPAPAKP-----EQSDSD-----SEGAAGPAGAPRT 96
Db 152 VSVKPPRGRKPAEKPLPKPRGKPKPERPPSSSSSDSDSDSEVDRISEWKRRDEARRREL 211
QY 97 LVRRRR-----RRL-----LDGEPAPVPVYSGKVQSSNLIPDSSLLJK 136
Db 212 EARRRREOEELRLRLEQEKERRRRADRGEAE-----RSGSGSGDELREDDPEVK 266
QY 137 -----LCPSPED--EADLTNSGSSPSEDDALPSPGWRKLRKKEKEKK 181
Db 267 KGRKRGGRGPPSSDSEPEAELEAKKSAKKPOSSSTEPARKPQKE--KRVPEEKQ 324
QY 182 MEEFPDQDISPLPOPSSRNKSRKHTALQKLRVNRKLDLRSLCSPKQHSPALQSTDD 241
Db 325 Q-----AKPVKVERTKRSEGFMDRKVEKKE-----PSVEEK--LQKLHS 364
QY 242 EVVLVEGVPVLPQSSRLFTLKI-----RCRADLVRLPVRM--SEPLQNVVDHMANHLGVS 293
Db 365 EI-----KFALKVDSPPVKRCLNAALEELGTLTQVTSQILQKNVDVVA----- 405
QY 294 PNRILLFGESLSPTATPTSLK-----LGVADIIDCVVLASSSEATTSQELRLRVQCK 348
Db 406 -----TLKKIRRYKANKDVME-----KAAEVYTLKSRVLGP 437
QY 349 EKHQMLEISLSPSLKVLMSHYEAMGLSGHKLSPFFDGTGTLKSGKELPADLG 401
Db 438 KTEAVQVYKNA-----GMEKEKAEKLAGELAGEEAPQEK 474
```

```
RESULT 4
US-10-052-586-346
; Sequence 346, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
```

APPLICANT: Pan,James  
APPLICANT: Smith,Victoria  
APPLICANT: Watanebe,Colin K.  
APPLICANT: Wood,William I.  
APPLICANT: Zhang,Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P343OR1C1  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059266  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063120  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063121  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063486  
PRIOR FILING DATE: 1997-10-21  
PRIOR APPLICATION NUMBER: 60/063540  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063541  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063544  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063564  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063734  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063870  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/064103  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066120  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066466  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066772  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069335  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069425  
PRIOR FILING DATE: 1997-12-12  
PRIOR APPLICATION NUMBER: 60/069870  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: 60/068017  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333

PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086023  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086486  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087098  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087208  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088033  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088167  
PRIOR FILING DATE: 1998-06-05

```

; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088722
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088740
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088811
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088825
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908
```

```

Query Match      5.9%; Score 124.5; DB 12; Length 671;
Best Local Similarity 23.1%; Pred. No. 0.19;
Matches 93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;

Qy 8 RGRPRSGRGARRARGRCPRARQPARLPDPTVLVDLV-----SDSD 52
Db 213 RAR-RGPLGRKKK-----KAPASDSKADSDGAKPEVAVARASSSSSSSSDSD 267

Qy 53 EEVLE-----VADPVEVPVAPLAPAKP-----EQDSOD-----SEGAAGPAGAPRT 96
Db 268 VSVKPPGRGKPAEKPLPKPRGKPKPERPPSSSSSDSDSDVDVIRSEWKRDRARRREL 327

Qy 97 LVRRRR-----RL-----LDPGEAPVVPYSGKVOSLNLIPDNSSLK 136
Db 328 EARRRREQEELRLREKEKEKERRERADRGAE-----RSGGSSGDELREDDEPVK 392

Qy 137 -----LCSEPED--EADLTNSGSSPSEDDALPGSPWRKKLR----- 172
Db 383 KRGRKGRGCRPPSSDSEPAELERAKKSAKXPQSSSTEPARKGQKEKVRPEEKQA 442

Qy 173 --KKCKEKKMBEFP--DQDISPLPOSSNRKSKRTEALQKLR--EVNKRLO-----DLRS 224
Db 443 KPVKVERTKRSEGSMDRKRVEKKEPSVEEK-----LQKLHSEIKFALKVDSPDKR 495

Qy 225 CLS-----PKHQSPALQSTDDVEVLVE-----GPVLPOSSRLFTLKIRCRADLV 270
Db 496 CLNALBELGLQVTSQILQKNQDVATLKIRRYKANKOVNKAEEVIT-----R 545
```

```

Qy 271 LPVRMSEPLQNVVDHMANHLGVSPNRIILLLFGESELSPTATP 312
Db 546 LKSRVLGPKIEAVQKV-NKAGMEKEKABEKLAGEELAGEEAP 586

RESULT 5
US-09-823-356-9
; Sequence 9, Application US/09823356
; Patent No. US20010025098A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN MEMBRANE SPANNING PROTEINS
; FILE REFERENCE: PP-0489-1 CON
; CURRENT APPLICATION NUMBER: US/09/823,356
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/039,307
; PRIOR FILING DATE: 1998 March 13
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010025098A1 1794154
US-09-823-356-9
```

```

Query Match      5.9%; Score 123; DB 10; Length 950;
Best Local Similarity 19.9%; Pred. No. 0.38;
Matches 108; Conservative 74; Mismatches 203; Indels 158; Gaps 23;

Qy 3 EPLRGGRPRSGRGARRARGRCPRARQSP-----ARLPDPTVLVDLVSDSDDEVLEV 58
Db 314 DPRRAQVPPPTRESR-VKVLRLAATAPALPOPPSTPRATTLPPAPATTVTRTSRAVTV 372

Qy 59 ADPV-----EVPVARLPA-----PAKPEQSDS----- 80
Db 373 ARPMTTTPPTTQRTWTPSPSHRPTTTEVITARRPSVSENLYPPSRKDDQHRERPTTTR 432

Qy 81 SDSEGAAGPAGAPRTLVRRRRRRLDPG-----EAPVVPYSGKVQSSL 125
Db 433 PSKATSLSEFTNAPTITISEPSTRAAGPGFRPNRMDRREHGHDRPNVVP---GPPKPAK 489

Qy 126 NLIPDNSSLKLCPSPEDEADLTNSGSSPSDDALPGSPWRK-----KLKKCKEK 177
Db 490 EKPPKKAQDKILSEYBEKYDLSRPTASQLEDELQVGNVPLKKAKESKKHEKLEKEPE 549

Qy 178 EKKME-BFPDQDISPLPOPSSNRKSKHTEALQKLREVNKLQD----- 221
Db 550 KKKKMKMNADKLLKSEKQMKKSEKSKKQEKESKKKGGKTEQDQYQKPTNKHFTQSPK 609

Qy 222 -----LRSLSPKQHSQSPALQSTDD-----EVLVEGVPVLPO 253
Db 610 KSVADLIGSFEGRLLLLITAPKAENMVMVQQRDEYLESFCKMATRKISVITFGPV--- 666

Qy 254 SSRLLFTLKIRCRADLVRL-----PVRM--SEPL--QNVVDHMANHLGVSPNRIILLFGESE 305
Db 667 --NNSTMKI-----DHFDLDNEKPMRVDDDELVDQRLISLSEAKYGMTYNDFMVLTQVD 720

Qy 306 LSPT---ATPSTKLGVADIICVVVLASSSEATETSQELRLRVQGEK-KHOMLEISLSP- 360
Db 721 LRVQYVEVPIIMK-SVFDLID-----TFQSRIDKMEKQKKEGIVCKEDKKQSLFNLSRF 775
```

QY 361 -----DPLKVMGSHYEBAMGLSGHLSF---FPDGTKLKSLPADLGLSGDILI 408  
DB 776 RWRRLVLVSAIBNDEMDWASQOLSLASGQACNFGRLHITILKLG-----VGEVGVGL 829  
QY 409 EVM 411  
DB 830 ELP 832

## RESULT 6

US-09-738-877-3  
; Sequence 3, Application US/09738877  
; Patent No. US20020015970A1  
; GENERAL INFORMATION:  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan  
; APPLICANT: Weiss, Stephen J.  
; APPLICANT: Glynn, Richard  
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET  
; FILE REFERENCE: A-69806/DJB/JJD  
; CURRENT APPLICATION NUMBER: US/09/738,877  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: US 09/637,977  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: US 60/148,425  
; PRIOR FILING DATE: 1999-08-11  
; PRIOR APPLICATION NUMBER: PCT/US 00/22061  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO: 3  
; LENGTH: 1781  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-877-3

Query Match 5.8%; Score 122.5; DB 10; Length 1781;  
Best Local Similarity 22.0%; Pred. No. 0.92;  
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;

QY 40 PPTV-LVDIVSDSEVLEVADPVEVPVARLPAPAK---PEODSDSDEGAEGPAGAPR 95  
DB 192 PPTVOLLTVKKEGEGAGAGADHOPSLGAGEASKESPEKOSTEKPE-----240  
QY 96 TLVRRRRRLDPCGAPVVPVYSGKVQSSLNLPDSSLLKLCPSPEPD--EADLTNSGS 153  
DB 241 TLKRQSHAEISP-----PAESQA-----VEECKEKEKEKEPEKSAE 281  
QY 154 SPSEDDALPSGSP-----WRKK--LRKKEKE-----EKKMEPEPDODISPLPQ 196  
DB 282 SPTSIVTSTGTFKFTFOGAGNRKKTSPFKPKEDVEASEKKKEQEPK-----VDT 336  
QY 197 SSRNRSRKHTALQKLRVYNKRLQDLRSLSPK---QHOSPALQSTDEVDVLEGP--- 249  
DB 337 EEDGAEVASEKLTASSEQNHQ--EPASAHBEPRLSAEYKVELPS--EEGVSSGSGSEK 394  
QY 250 VLPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDMANHLGVSPNRILLFGESLSFT 309  
DB 395 PAPLATEVFEDEKIEVHOEVAEVAHVSTVEERTEQ-----KTEVEERT 437  
QY 310 A--TSTTKLGVADIIDCVLASSSEATETSOELRLR---VQKKEGQMLLETISLPSDL 364  
DB 438 AGSVAEAEIVGND-----AEPOEAEPAKELVKIKETCVSGEDTQCAD--LSDEKV 487  
QY 365 -----KVLNMGHYEBAMGLSGHLSFPDGT---KLSGKELPADLG---LESGLLI 408  
DB 488 LSKPEGVVSEVEMLSGGERMKVQSPKLKLTFTSTGLTKLGGKQKQKGGGDBESGERT 547  
QY 409 EV 410  
DB 548 QV 549

## RESULT 7

US-09-765-272-218  
; Sequence 218, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,272  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 218:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 565 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 218:  
US-09-765-272-218

Query Match 5.7%; Score 120.5; DB 10; Length 565;  
Best Local Similarity 20.9%; Pred. No. 0.29;  
Matches 81; Conservative 66; Mismatches 156; Indels 85; Gaps 19;

QY 29 PRARQSP-----ARLIPTVLVDIVSDSEVLEVADPVEV--PVARLP---APAKPE-Q 77  
DB 34 PFTESPEEPEKSEVKKPDDTLPKVEEGKEDSABAPVVEVGEVESKPEEKVAVKPEQ 93  
QY 78 DSDSDEGAEGPAGAPPTLVRRRRRLDPCGAPVVPVYSGKVQSS--LNLIPDSSSL 134  
DB 94 PSDKRAESKVEQNAEP--VAREDEKAPVEPEKQPEAEBAVEEKAIVETPKQESTDTYAE 152  
QY 135 LKLCP-----SEPEDEADLTNSGS--SPSEDDALPSGSPWRKK 170  
DB 153 ETVPEKEETVQSIQPKVETPAVEKQTEPEEPKVEQAGFVAPREDEQAFT--APVEPE 211  
QY 171 LRKKEKEKEKEEPPDDDISPLPOPSSRNRSRKHTALQKLRVYNKRLQDLRSLSPKQ 230  
DB 212 KQPEVPEEKAAVE-----TPRPEDKINGITGKEPVDR--SEINNOI--DRASSVSPD 261  
QY 231 HQSPALQSTDEVDVLEGPVLPQSSRLFTLKIRCRADLVRLPVRMSE--PLQNVVD--- 284  
DB 262 YSTASVNL-----GPVLETAKGVYA-----SEPVKQEVVNSFNKLTALDALNV 307  
QY 285 ---HMANHLGVSPNRILLFGESLSPTATPSTKLGVADIIDCVLASSSEATETSOEL 341  
DB 308 DKTELNTIADAKTKVKEHYSDRSWONLQTEVTYKAEKVAANTD---AKOSEVNEAVEKL 363

QY 342 RLVOGKERHQMLEISLSPDPLKVLMS 369  
Db 364 TATIE-----KLVELS---EXPILTITS 383

## RESULT 8

US-09-931-087A-1  
; Sequence 1, Application US/09931087A  
; Patent No. US20020147322A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; TITLE OF INVENTION: NSP Molecules  
; FILE REFERENCE: P1223RIE  
; CURRENT APPLICATION NUMBER: US/09/931,087A  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/367,206  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/082,767  
; PRIOR FILING DATE: 1998-04-23  
; PRIOR APPLICATION NUMBER: US 60/113,296  
; PRIOR FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 35  
; SEQ ID NO 1  
; LENGTH: 576  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-931-087A-1

Query Match 5.6%; Score 118; DB 10; Length 576;

Best Local Similarity 21.8%; Pred. No. 0.45;  
Matches 90; Conservative 49; Mismatches 148; Indels 126; Gaps 18;

QY 3 EPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLSDSDSEVLEVADPV 62  
Db 137 EPLRAR-----KWSNQPADLAHMGSRSDPAGMEASTMPISALPRTS-----SDPV 183  
QY 63 EYPVARLPAPAKPEQDS--DSDSEGAAGPAGAPRTLVRRRRRRLDPGEAPVVPVYSGK 120  
Db 184 ---LLKAPAPLGTVDASLRASDQLOKAKTPKPR-----PSFELPDASERP--PTY--- 231  
QY 121 VOSSLNLIPDNSSLLKCPSEPEDEADLTNSGSSPSEDALPSGSPWRKKLRKCKEKEK 180  
Db 232 -----CELVRPVPSVQ-----GTSQSCPEPE-APWWE--AEDEEEN 268  
QY 181 KMEFPDQDISPLPQSPSRNKRKHTKALQKLVNKRLODLRSLSPKQHO--SPALQS 238  
Db 269 RCFTRPQAEISFCP-----HDAPSCLLGPONRPLEPOVLH 303  
QY 239 TDDEVVLVEGPVL---POSSRLFTLKIRCA-----DLVRLPVRM 275  
Db 304 T-----LRGLFLEHPGSTALHLLVDCOATGLLVTRDQGNMGVSSGLELLTLPHG 357  
QY 276 SEPLQNVVDHMANHL-----GVSPNRILLFGESELSPTATPSTL--KLGVADII 323  
Db 358 HRLLELLEHQTALAGALAVLGCSPLEERAAALRGVLALALRPGNAGDPLGLAAYM 417  
QY 324 DCVVLASSEATTSQELRLRVQKQHEQMLISLSPDPLKVLMSHYEAMG 376  
Db 418 GALLMPQVSRLEHTWRQLR-----RSHTAALAFEOELKPLMRALDEGAG 462

## RESULT 9

US-09-931-087A-21  
; Sequence 21, Application US/09931087A  
; Patent No. US20020147322A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; TITLE OF INVENTION: NSP Molecules  
; FILE REFERENCE: P1223RIE  
; CURRENT APPLICATION NUMBER: US/09/931,087A  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/367,206

; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/082,767  
; PRIOR FILING DATE: 1998-04-23  
; PRIOR APPLICATION NUMBER: US 60/113,296  
; PRIOR FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 35  
; SEQ ID NO 21  
; LENGTH: 576  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Mutation of SEQ ID NO:1  
US-09-931-087A-21

Query Match 5.6%; Score 118; DB 10; Length 576;

Best Local Similarity 21.8%; Pred. No. 0.45;  
Matches 90; Conservative 49; Mismatches 148; Indels 126; Gaps 18;

QY 3 EPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLSDSDSEVLEVADPV 62  
Db 137 EPLRAR-----KWSNQPADLAHMGSRSDPAGMEASTMPISALPRTS-----SDPV 183  
QY 63 EYPVARLPAPAKPEQDS--DSDSEGAAGPAGAPRTLVRRRRRRLDPGEAPVVPVYSGK 120  
Db 184 ---LLKAPAPLGTVDASLRASDQLOKAKTPKPR-----PSFELPDASERP--PTY--- 231  
QY 121 VOSSLNLIPDNSSLLKCPSEPEDEADLTNSGSSPSEDALPSGSPWRKKLRKCKEKEK 180  
Db 232 -----CELVRPVPSVQ-----GTSQSCPEPE-APWWE--AEDEEEN 268  
QY 181 KMEFPDQDISPLPQSPSRNKRKHTKALQKLVNKRLODLRSLSPKQHO--SPALQS 238  
Db 269 RCFTRPQAEISFCP-----HDAPSCLLGPONRPLEPOVLH 303  
QY 239 TDDEVVLVEGPVL---POSSRLFTLKIRCA-----DLVRLPVRM 275  
Db 304 T-----LRGLFLEHPGSTALHLLVDCOATGLLVTRDQGNMGVSSGLELLTLPHG 357  
QY 276 SEPLQNVVDHMANHL-----GVSPNRILLFGESELSPTATPSTL--KLGVADII 323  
Db 358 HRLLELLEHQTALAGALAVLGCSPLEERAAALRGVLALALRPGNAGDPLGLAAYM 417  
QY 324 DCVVLASSEATTSQELRLRVQKQHEQMLISLSPDPLKVLMSHYEAMG 376  
Db 418 GALLMPQVSRLEHTWRQLR-----RSHTAALAFEOELKPLMRALDEGAG 462

## RESULT 10

US-09-931-087A-22  
; Sequence 22, Application US/09931087A  
; Patent No. US20020147322A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; TITLE OF INVENTION: NSP Molecules  
; FILE REFERENCE: P1223RIE  
; CURRENT APPLICATION NUMBER: US/09/931,087A  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/367,206  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/082,767  
; PRIOR FILING DATE: 1998-04-23  
; PRIOR APPLICATION NUMBER: US 60/113,296  
; PRIOR FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 35  
; SEQ ID NO 22  
; LENGTH: 576  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Mutation of SEQ ID NO:1  
US-09-931-087A-22

Query Match 5.6%; Score 118; DB 10; Length 576;



Best Local Similarity 21.8%; Pred. No. 0.45; Indels 126; Gaps 18;  
Matches 90; Conservative 49; Mismatches 148;  
QY 3 EPLRGGRSRRGRRGARRGRCPPAROSPARLIPDTVLVDVSDSEVLVADPV 62  
DB 137 EPLRAR-----KMSNSGPADLAHGRSREDPAGMEASTW1SALPRS-----SDPV 183  
QY 63 EYVVARLPAPAKPEODS--DSDEGAIEGPAGAPRTLVRRRRRLIDPGEAVVPVYSCK 120  
DB 184 ---LTKAPAPLCTVADSLRASDGOAKAPTKPRT-----PSFELPDASERP--PTY--- 231  
QY 121 VQSSUNLIPDSSSLKLCSPSEPEDEADLTNCGSSPSEDALSGSGWKKLAKCKEK 180  
DB 232 ---CELVRPVSVQ-----GTSPOSCPEPE-APWWE--AEEDEEEN 268  
QY 181 KMEEFPPDODISPLPPSSRNKSRKHTALQKLEVNKRLODRSCSPPOHO--SPALOS 238  
DB 269 RCFTRPQALISFCP-----HDAPSCLLGQNPRLPEQVUH 303  
QY 239 TDEVVVLVEGPVL---POSSRLFTLKIRCA-----DLVRLPVRM 275  
DB 304 T-----LRLGLEHHPGSTALLHLVDCQATGLGTRDRGNMGVSSGLELTLPHGH 357  
QY 276 SEPLQNVVDHMANHL-----GVSPNRILLFGESELSPTATPSTL--KLGVADI 323  
DB 358 HURLLELHQTALAGALAVLGCSPLEERAAALRGVELALALRPGAAGDLPGLAAVM 417  
QY 324 DCVVLASSEATETSOELRLRVQGEKQHMLEISLSPDSPKLVMSHYEAMG 376  
DB 418 GALLMPQVSRLEHTWRQLR-----RSHTEAALAFEOELKPLMRALDEGAG 462  
RESULT 11  
US-09-931-087A-23  
Sequence 23, Application US/09931087A  
Patent No. US20020147322A1  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: NSP Molecules  
FILE REFERENCE: P1223RLE  
CURRENT APPLICATION NUMBER: US/09/931,087A  
CURRENT FILING DATE: 2001-08-15  
PRIOR APPLICATION NUMBER: 09/367,206  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: US 60/082,767  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: US 60/113,296  
PRIOR FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 35  
SEQ ID NO 23  
LENGTH: 576  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Mutation of SEQ ID NO:1  
US-09-931-087A-23  
Query Match 5.5%; Score 115; DB 10; Length 576;  
Best Local Similarity 21.1%; Pred. No. 0.74; Indels 134; Gaps 18;  
Matches 88; Conservative 52; Mismatches 143;  
QY 3 EPLRGGRSRRGRRGARRGRCPPAROSPARLIPDTVLVDVSDSEVLVADPV 62  
DB 137 EPLRAR-----KMSNSGPADLAHGRSREDPAGMEASTW1SALPRS-----SDPV 183  
QY 63 EYVVARLPAPAKPEODS--DSDEGAIEGPAGAPRTLVRRRRRLIDPGEAVVPVYSCK 120  
DB 184 ---LTKAPAPLCTVADSLRASDGOAKAPTKPRT-----PSFE-- 220  
QY 121 VQSSUNLIPDSSSLKLCSPSEPEDEADLTNCGSSPSEDALSGSGWKKLAKCKEK 176  
DB 221 ---LPPASERPPTECELVRPVSVQ-----GTSPOSCPEPE-APWWE--AEEDE 264

QY 177 KEKKMEFPDODISPLPPSSRNKSRKHTALQKLEVNKRLODRCLSPKHO--SP 234  
DB 265 EENRCFTRPQALISFCP-----HDAPSCLLGQNPRLPEQ 299  
QY 235 ALQSTDEVVVLVEGPVL---POSSRLFTLKIRCA-----DLVRL 271  
DB 300 QVLT-----LRLGLEHHPGSTALLHLVDCQATGLGTRDRGNMGVSSGLELTL 353  
QY 272 PYRMSEPLQNVVDHMANHL-----GVSPNRILLFGESELSPTATPSTL--KLGV 319  
DB 354 PHGHRLLELHQTALAGALAVLGCSPLEERAAALRGVELALALRPGAAGDLPGL 413  
QY 320 ADIDCVVLASSEATETSOELRLRVQGEKQHMLEISLSPDSPKLVMSHYEAMG 376  
DB 414 AAVMGALLMPQVSRLEHTWRQLR-----RSHTEAALAFEOELKPLMRALDEGAG 462  
RESULT 12  
US-09-864-761-36307  
Sequence 36307, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rankel, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmicea-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
SEQ ID NO 36307  
LENGTH: 92  
TYPE: PRT  
ORGANISM: Homo sapiens

```

;
; FEATURE:
; OTHER INFORMATION: MAP TO ALO31133.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN SWISS3POT HIT: P55855, EVALUE 9.00e-40
; OTHER INFORMATION: EST_HUMAN HIT: BF103822.1, EVALUE 8.00e-39
;
US-09-864-761-36307

Query Match 5.5%; Score 114.5; DB 10; Length 92;
Best Local Similarity 33.8%; Pred. No. 0.075;
Matches 26; Conservative 17; Mismatches 33; Indels 1; Gaps 1;

QY 335 TETSQELRLRVQGEKHQMLEISLSPDPLKVLMSHYEEAMGLSGHKLFFDFDGTKLXGK 394
DB 12 TENNHNLKVAQOD-GSVVQFKIKQTPLSKLMKAYCEPRGLSVKQIRFRFGQPISGT 70
QY 395 ELPADLGLESGDLIEVW 411
DB 71 DKPAQLEMEDEDTIDVF 87

RESULT 13
US-10-160-865-14
; Sequence 14, Application US/10160865
; Patent No. US20020169139A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
; FILE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
; FILE REFERENCE: 05433/038001
; CURRENT APPLICATION NUMBER: US/10/160,865
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US/09/134,250
; PRIOR FILING DATE: 1998-08-14
; PRIOR APPLICATION NUMBER: US 08/795,868
; PRIOR FILING DATE: 1997-02-06
; PRIOR APPLICATION NUMBER: US 08/494,577
; PRIOR FILING DATE: 1995-06-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-10-160-865-14

Query Match 5.4%; Score 113.5; DB 9; Length 661;
Best Local Similarity 20.4%; Pred. No. 1.1;
Matches 93; Conservative 55; Mismatches 121; Indels 188; Gaps 22;

QY 24 ARGCRPARQSPARLIPDTVLVLVDSDE--EVLEVADPVEVPV-----ARLPAPA 73
DB 85 ALGRSPLRVAGSRIL-----DKLOPFERRRSLERSDPPAPLPWPLRKARSLEQP 138
QY 74 KPEQDSDSDEGAAGPAGAPRTLVRRR-----RRLLDGCEAPVVPVYS-----G 119
DB 139 KSERGAPWGTGPGSQEELRAGPVAERRLRFQKAAASLDBTRORSFASDLERFAQELG 198
QY 120 KVQSSLN---LIPNSSL---LKLCPSEPE-DEADLTNNGSSPSSEDDAL----- 161
DB 199 RIRSTREELVRSHESLRATLQRAPSPREGEPLFSRSTPTKTSRAVSPAAQAQPPSPS 258
QY 162 ----PSGSPMVKLRKXC-----EKEKKMEEPF-----DQISPLPOPS----- 197
```

```

DB 259 SAEKPGDEPRSRGPAGRTPEGPGPOQEVRRRDQFPLTRSRALQECRSPVPPPAADPP 318
QY 198 -SRNKS---RKHTALQKLR-----EVNKRLO----- 220
DB 319 EARTKAPGRKREPPAQAQVRFPLPWATPGLEGAAVPTLEKNRAGPEAEKRLRRRGPEDGP 378
QY 221 ---DLRSLSPKHQ-----SPALQSTDDDEVL-----VSGPVLPOSSRLFTLKIRCRA 266
DB 379 WGPWDRRGARSQGGKRRARPTSPLESSDDSVSAGESPLEAPV----- 422
QY 267 DLVRLPVRMSEPLQVVDHMANHLGVSPNRILLFGESELSPTATPSTLKLGVADIIDCV 326
DB 423 -----FEPLQNVV-----VAP-----GADVLLKCI 443
QY 327 VLAS-----SSEATETSOELRLRVQGEKHQML 354
DB 444 ITANPPQVQSMHKGSLRSEGRLLLRAG-ERHTLL 479

RESULT 14
US-10-028-072-160
; Sequence 160, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
```

PRIOR APPLICATION NUMBER:	60/062814
PRIOR FILING DATE:	1997-10-24
PRIOR APPLICATION NUMBER:	60/062816
PRIOR FILING DATE:	1997-10-24
PRIOR APPLICATION NUMBER:	60/063045
PRIOR FILING DATE:	1997-10-24
PRIOR APPLICATION NUMBER:	60/063082
PRIOR FILING DATE:	1997-10-31
PRIOR APPLICATION NUMBER:	60/063127
PRIOR FILING DATE:	1997-10-24
PRIOR APPLICATION NUMBER:	60/063327
PRIOR FILING DATE:	1997-10-27
PRIOR APPLICATION NUMBER:	60/063329
PRIOR FILING DATE:	1997-10-27
PRIOR APPLICATION NUMBER:	60/063550
PRIOR FILING DATE:	1997-10-28
PRIOR APPLICATION NUMBER:	60/063561
PRIOR FILING DATE:	1997-10-28
PRIOR APPLICATION NUMBER:	60/063704
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/063733
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/063755
PRIOR FILING DATE:	1997-10-17
PRIOR APPLICATION NUMBER:	60/063755
PRIOR FILING DATE:	1997-10-23
PRIOR APPLICATION NUMBER:	60/064809
PRIOR FILING DATE:	1997-11-07
PRIOR APPLICATION NUMBER:	60/065186
PRIOR FILING DATE:	1997-11-12
PRIOR APPLICATION NUMBER:	60/065846
PRIOR FILING DATE:	1997-11-17
PRIOR APPLICATION NUMBER:	60/066364
PRIOR FILING DATE:	1997-11-21
PRIOR APPLICATION NUMBER:	60/066453
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/066511
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/066770
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/069334
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069278
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069334
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069694
PRIOR FILING DATE:	1997-12-16
PRIOR APPLICATION NUMBER:	60/072320
PRIOR FILING DATE:	1998-01-23
PRIOR APPLICATION NUMBER:	60/073612
PRIOR FILING DATE:	1998-02-04
PRIOR APPLICATION NUMBER:	60/074086
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/074092
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/077791
PRIOR FILING DATE:	1998-03-12
PRIOR APPLICATION NUMBER:	60/078910
PRIOR FILING DATE:	1998-03-20
PRIOR APPLICATION NUMBER:	60/079294
PRIOR FILING DATE:	1998-03-25
PRIOR APPLICATION NUMBER:	60/079663
PRIOR FILING DATE:	1998-02-27
PRIOR APPLICATION NUMBER:	60/079728
PRIOR FILING DATE:	1998-03-27
PRIOR APPLICATION NUMBER:	60/080165
PRIOR FILING DATE:	1998-03-31
PRIOR APPLICATION NUMBER:	60/081203

PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081229
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081695
PRIOR FILING DATE:	1998-04-14
PRIOR APPLICATION NUMBER:	60/081871
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/081818
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/082999
PRIOR FILING DATE:	1998-04-24
PRIOR APPLICATION NUMBER:	60/083322
PRIOR FILING DATE:	1998-04-28
PRIOR APPLICATION NUMBER:	60/083545
PRIOR FILING DATE:	1998-04-29
PRIOR APPLICATION NUMBER:	60/084600
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/084662
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/085149
PRIOR FILING DATE:	1998-05-12
PRIOR APPLICATION NUMBER:	60/085322
PRIOR FILING DATE:	1998-05-13
PRIOR APPLICATION NUMBER:	60/085338
PRIOR FILING DATE:	1998-05-13
PRIOR APPLICATION NUMBER:	60/085333
PRIOR FILING DATE:	1998-05-13
PRIOR APPLICATION NUMBER:	60/085579
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/085697
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/085704
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/086414
PRIOR FILING DATE:	1998-05-22
PRIOR APPLICATION NUMBER:	60/086430
PRIOR FILING DATE:	1998-05-22
PRIOR APPLICATION NUMBER:	60/087106
PRIOR FILING DATE:	1998-05-26
PRIOR APPLICATION NUMBER:	60/088028
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088730
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088741
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088810
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088858
PRIOR FILING DATE:	19/98-06-11
PRIOR APPLICATION NUMBER:	60/089352
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089599
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089907
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089947
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/090349
PRIOR FILING DATE:	1998-06-23
PRIOR APPLICATION NUMBER:	60/090429
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090445
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090588
PRIOR FILING DATE:	1998-06-26
PRIOR APPLICATION NUMBER:	60/091366
PRIOR FILING DATE:	1998-07-01
PRIOR APPLICATION NUMBER:	60/091519
PRIOR FILING DATE:	1998-07-02



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 10:25:28 ; Search time 22 Seconds

(without alignments)  
1800.336 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099  
Sequence: 1 MAEPLRGSRGSRGARR.....GKELPADLQESGLIEVWG 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.5	7.4	1217	2 T00270	hypothetical prote
2	145.5	6.9	1072	1 A37221	neurofilament trip
3	140	6.7	1634	2 T26517	hypothetical prote
4	137.5	6.6	854	2 S02003	neurofilament trip
5	134.5	6.4	587	1 WMBETE	65k early nonstruc
6	133	6.3	971	2 T19431	hypothetical prote
7	132.5	6.3	792	2 T29187	hypothetical prote
8	132	6.3	586	1 WMBEDE	65k early nonstruc
9	128.5	6.1	635	2 T09648	nucleolin homolog
10	128.5	6.1	734	2 B42680	nucleolin-cytoplas
11	128.5	6.1	742	2 D87719	protein R119.4 (m
12	128	6.1	303	1 Z8BP22	gene 8 protein - p
13	127	6.1	1365	2 S14871	suppressor two of
14	126	6.0	856	2 T16543	hypothetical prote
15	124.5	5.9	1684	2 T00057	gravin - human
16	124.5	5.9	1791	2 T02345	hypothetical prote
17	122	5.8	1082	2 T15269	hypothetical prote
18	120.5	5.7	1881	2 H95076	zinc metalloprotei
19	120	5.7	669	2 JCS662	hepatoma-derived g
20	119	5.7	450	2 T08701	hypothetical prote
21	119	5.7	765	2 T49592	neurofilament trip
22	119	5.7	1603	2 S17983	gene posterior sex
23	119	5.7	1611	2 T38236	hypothetical prote
24	118.5	5.6	508	2 T04605	hypothetical prote
25	118	5.6	517	2 T12536	hypothetical prote
26	117.5	5.6	1110	2 T15116	NF-180 - sea lamp
27	117.5	5.6	2187	2 T30826	nascent polypeptid
28	117	5.6	553	2 C75318	hypothetical prote
29	116.5	5.6	714	2 T22454	hypothetical prote

30	116.5	5.6	955	2 S52959	male-specific lecth
31	116.5	5.6	974	2 A40580	lodestar maternal-
32	116	5.5	672	2 H86169	hypothetical prote
33	116	5.5	1043	2 A56037	DNA-binding protei
34	115.5	5.5	872	2 S62061	SCD5 protein - yea
35	115.5	5.5	1280	2 T00385	hypothetical prote
36	115	5.5	382	2 T14336	Rad23 protein, iso
37	114	5.4	891	2 G84693	probable proline-r
38	114	5.4	1274	2 T16251	hypothetical prote
39	113.5	5.4	1119	2 AC0045	probable membrane
40	113.5	5.4	2361	2 T25752	hypothetical prote
41	113	5.4	611	2 T06458	nucleolin homolog
42	113	5.4	776	2 T20738	hypothetical prote
43	113	5.4	1083	2 C88854	protein F1A10.3 f
44	112.5	5.4	95	2 JC4760	SM73 protein - hum
45	112.5	5.4	675	2 T03744	myod protein inhib

ALIGNMENTS

RESULT 1

T00270  
hypothetical protein KIAA0596 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000

C:Accession: T00270

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complet

A:Reference number: Z14086; MUID:98290545; PMID:9628581

A:Accession: T00270

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1217 <NAG>

A:Cross-references: EMBL:AB011168; NID:G3043715; PIDN:BA25522.1; PID:G3043716

A:Experimental source: Brain

C:Genetics:

A:Note: KIAA0596

Query Match	7.4%	Score 154.5;	DB 2;	Length 1217;
Best Local Similarity	23.5%	Pred. No. 0.11;		
Matches 105;	Conservative 48;	Mismatches 164;	Indels 129;	Gaps 21;
QY	6	RGRGSRSGRGARARGRCPRARSPARLIPDTVLVDLVSDSEVLEADVPEVP	65	
DB	455	KRGKQGGPSSPQFASG---PVRHQPSPMLSPGA---LSSDSKGEDEGTETELP	506	
QY	66	-----VARLPAPAKPEODSDSEGAAE-----GPAGAPRTLVRRRRRL	106	
DB	507	ALPVLASTKKALASVSPALPRSLSHWEMRAQESVFLDPAPANP---RRGRWV	563	
QY	107	DPEAPVVPVYSGVQSSLNIPNSSILKLCPS--EP-EDPADITNSGSSPEDDALPS	163	
DB	564	QPG-----VELSVRSMFLRQLETFLAPSLQDPSQSLAIIIPSGPRKHGQELLET	612	
QY	164	G-----SPWKKLRKCC-----EKKE-----KKMEEPDDISLPPPS---SNK	201	
DB	613	SLTONEKPPRPASQSPSYHIIIRLSQEGVFAQDLEPAIDEGIVYPPDNPPTMDT	672	
QY	202	SRKHTALQK-----LREVNKRLOD-----LRSCSPKXOHSPALOSTDDEVVL	245	
DB	673	SEFQVQAPARETLGRVYVPGSSSEKHSKSDASCVDYSSSCLSSEPH--PTDSESTETLS	730	
QY	246	VEGVLPQSSRLFTLKIRCRADLVRLVPRMSEPIQNVVDHMANHLGVSPNILLPGESE	305	
DB	731	VDG-----ISSDLEPAGDEGEDEEEEGGMGP-----YGIQE	762	
QY	306	LSPATP-----STLKLGVADIIDCVVLASSSATSTQELRLRVQGXKXHQGLE	355	
DB	763	GSP-QTPDQEOFLKHQFETLLSGAAPGAP-VQVPERSESRISIRFLLVQTRPLREP--	818	
QY	356	ISLSPDSPLKVLMSHYEFAMGLSGHK	381	



A:Reference number: S02003, MUID:89065087, PMID:3143606  
 A:Accession: S02003  
 A:Molecule type: mRNA  
 A:Residues: 1-854 <BRE>  
 A:Cross-references: EMBL:X13804; NID:957828; PIDN:CA32038.1; PID:957829  
 C:Superfamily: neurofilament triplet H protein  
 C:Keywords: coiled coil

Query Match 6.6%; Score 137.5; DB 2; Length 854;  
 Best Local Similarity 21.9%; Pred. No. 0.77;  
 Matches 94; Conservative 62; Mismatches 168; Indels 105; Gaps 20;

QY 4 PLRGRGPRGRCGARRARARGRCPARQSPARL-IDTVLVLDVSDSEV----- 55  
 DB 372 PVEAKSPAEAKSPASVSKSPG-EAKSPAEAKSPAEVSPATVSKPVEAKSPAEVSPVTK 430  
 QY 56 --LEVADPVEVVARLPAPAKPEODSDSDSGAEGPAGARTIVRRRRRLDPCGAPV 113  
 DB 431 SPAEAKSEVY-----KSPASVSPSEAKSPAGAKSPAEAKSPVAKSPAEAKSPAGAK- 484  
 QY 114 VPVYSGVQSSLNLPDSSSLKLCPSPEDEADLTN-----SGSPSEDDA---LP 162  
 DB 485 -PPEAKSPAEAKSPAEAKS-----PAKSPAEAKSPVEVSKSPAEKSPVEGAKSLAE 538  
 QY 163 SGSPWRKKLRKKCEKEKMEEPDQDISPL-POPSSRNKR---KHTALQKLRVNVK 217  
 DB 539 AKSPEKAKSPVKEIKPPAEVSKPEKAKSPKKEAKSPKAKTLDVKSPKAPKPEAK 598  
 QY 218 RLQDLRGLSKQKOSPALQSTDEEVLVVEGVLPOSSRLTLTKRCADLVRL--VR 274  
 DB 599 RPADIR---SEOVKSPAEKAKS-----PEKEETKVPKPEEVKSEVVEVK 645  
 QY 275 MSEPLQNVDMANHLGVSPRILLPGESLSPATPSTLKLGVADIDCVVLASSESA 334  
 DB 646 AKEP-----PKV-----EEKTP-ATPKT-----VKESKDEA 674  
 QY 335 TETSOELRLRVQGEKQKQMLISLSPDPLVLMSHYEAMGLSGHKLFFPDGTLKSGK 394  
 DB 675 PKEAQ---KPKAEKEPLTE--KPKDSPGEAKKEAKEKAAABE-----E 715  
 QY 395 ELPADLGLF 403  
 DB 716 ETPAKLGK 724

## RESULT 5

WMBETE

65K early nonstructural protein - human cytomegalovirus (strain Towne)

C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C&gt;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 16-Jul-1999

C:Accession: A41808

R:He, Y.S., Xu, L., Huang, E.S.

J. Virol. 66, 1098-1108, 1992

A&gt;Title: Characterization of human cytomegalovirus UL84 early gene and identification of

A:Reference number: A41808, MUID:92114132, PMID:1130892

A:Accession: A41808

A:Molecule type: mRNA

A:Residues: 1-587 &lt;HEY&gt;

A:Cross-references: GB:M81432; NID:9330541; PIDN:AAA45947.1; PID:9330542

C:Superfamily: human cytomegalovirus 65K early nonstructural protein

C:Keywords: leucine zipper; nonstructural protein

F:114-135/Region: leucine zipper motif

F:325-373/Region: leucine zipper motif

Query Match 6.4%; Score 134.5; DB 1; Length 587;  
 Best Local Similarity 24.4%; Pred. No. 0.74;  
 Matches 105; Conservative 58; Mismatches 159; Indels 109; Gaps 22;

QY 6 RGRPRGR-----GRRGARRAGRCRRARQS-----PARLIPDVLVDL--VSDSEV 55  
 DB 11 RARRPRARRGGGGVGSNSRRS-GKCRRRARRALSAPPLTFLATTTTTTMMGVASTDDSD 69

QY 56 LEVADPVEVVARLPAPAKPEODSDSDSGAEGPAGARTIVRRRRRLDPCGAPVP 115  
 DB 70 LLLKTPDEL-----DKHSGSPQTIL-----TLTDKHDIRQPR 101  
 QY 116 VYSGK---VQSSNLIPDN-----SSLKLCPSPEDEADLTNCGSSSEDDALPS 163  
 DB 102 VHRGYHLIQLHLDRPELRDPQIILLSTPLQGEANGESO---TADATQOEETAA-- 156  
 QY 164 GSPWRKKLRKKCEKEKMEEPDQDISPLPOPSSRNKRKHTMALQKLRVNVKLDLDR 223  
 DB 157 ---SHELKKEKEKEKEE--DED-----DRNDRER---GLCVANSNDSDVR 198  
 QY 224 SCLG--PKQHOSPALQSTDEEVLVVEGVLPOSSRLFTLTKRCADLVRLPYRKSEPLON 281  
 DB 199 PAFSLFPPAPCPCHILRSVIDQ-QLTRMAIVRLSLNLFALRI-ITPPLKRVPLRRKAHHT 256  
 QY 282 VV-DHMANHLGVSPRILLPGESLSPAT-PTLTKL-GVADIIDCVLASSSATETS 338  
 DB 257 ALHOCMALHL-----PELTPESTLDINNVTENASVADAESTDADLT 299  
 QY 339 QELRLRVQGEKQKQMLISLSPDPLKV-----LMSHYEAMGLSGHKLFFPDGTLKSG 393  
 DB 300 PTLTVRVNHAQWHRVEGGISGPRGLTISRISARLSETTAKTLGFS-----VFGRLDLP 353  
 QY 394 KELPADLGLFS 404  
 DB 354 NESPPDLTSS 364

## RESULT 6

hypochemical protein C25A1.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T19431

R:Mortimore, B.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19124

A:Accession: T19431

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-971 &lt;WIL&gt;

A:Cross-references: EMBL:Z81038; PIDN:CAB02755.1; GSPDB:GN00019; CESP:C25A1.10

A:Experimental source: clone C25A1

C:Genetics:

A:Gene: CESP:C25A1.10

A:Map position: 1

A:Insertions: 38/3; 92/3; 201/3; 919/3

C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein

Query Match 6.3%; Score 133; DB 2; Length 971;  
 Best Local Similarity 23.3%; Pred. No. 1.7;  
 Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;

QY 29 PRARQSPARLIPDVLVDLV-----DSDEVELEVADPVEVVARLPAPAKPEODSDSDSEG 85  
 DB 221 PAKPTPAKPTPKVVKKAESSSDSDDEKKPVAKP--APAKATPKPAKKAADSSSDSD 288  
 QY 86 ---AEGPAGAPRTLVRRRRLD-----PGEAPVPVYSGVQSSNLNLI 128  
 DB 289 DEAPAKTPAKAAPKVPVAKKAESSSDSDDEKKPAKPTPAKATPKPAKKAESSSDSD 348  
 QY 129 PDNSSLKLCPSF--PDEADLTNCGSSPSRDDALPSSSP--RKLTKRKCEKEKMEEF 185  
 DB 349 DEKKVPAPAKATPKVAKKAESSSDSDDEKKPAKPTPAKATPKVAKKAESSSDS 408  
 QY 186 PDQDISPLPOPSSRNKS-----RKHTALQKLRVNVKLDLDRSC 225  
 DB 409 SDDEKKVPAPKATPKATPKPAKKAADSSSDSDDEAPAKTPAKAAPKAKKA--SS 465  
 QY 226 LSPKQHOSPALQSTDEV 243  
 DB 466 SDSSDDEKPAKSTPAKI 483

## RESULT 7

T29187  
hypothetical protein C55C3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
A:Accession: T29187  
R:Woessne, J.; Stellyes, L.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid C55C3.  
A:Reference number: Z20585  
A:Accession: T29187  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-792 <WOE>  
A:Cross-references: EMBL:U53335; PIDN:AAA96170.1; GSPDB:GN00022; CESP:C55C3.3  
A:Experimental source: strain Bristol N2; clone C55C3  
C:Genetics:  
A:Map position: 4  
A:Introns: 17/1; 46/3; 139/2; 173/1; 204/3; 250/3; 266/3; 327/1; 362/3; 393/3; 43

Query Match 6.3%; Score 132.5; DB 2; Length 792;  
Best Local Similarity 20.9%; Pred. No. 1.4;  
Matches 90; Conservative 50; Mismatches 135; Indels 155; Gaps 19;  
QY 11 RSGRGARRARGRCRRAROSPARLIPDTVLVDLVSDSEVLEVADPVEVPVARLP 70  
DB 309 RRGGNQIKKKPAK---PLKRITP-----NNSDKLKKMKSP----- 343  
QY 71 APAKPQDSD-----SEGAAGPAGAPRTLVRRRR-----LLDPG 109  
DB 344 -PKPLQATSDILQPLPSNNAMTVPEKESANRSPQSKINLVNRGRTVNTDLKDCS 402  
QY 110 EAPVVPYSGKVOSSLNLPDNLKCPSEPEDEADLTNSGSSPSEDALPSGSPWK 169  
DB 403 EARAIP-----KC-DKLTDFQTSTNQSPA--DAISAVAP-TK 437  
QY 170 KLRKKEKEKKMEEPD-----QDI-----SPLQPSRNK----- 201  
DB 438 LTGSGQSEKQIEKMPNSFRELQDSKFKPAEAPIKPEERKSGTEQSTTEPTKLGSQ 497  
QY 202 SRKHTALQKLVNKRLOD--LRSCLSPKHQSPALQSTDD----- 241  
DB 498 SPSEKQIEQMPNSFRELQDSKFKPAPSPKSEKPAEQSLSESPSPVPGNNRRNSETSQVD 557  
QY 242 -----EVLVGGPVLQSSRLFTKIRCADLVRLPVRMSEPLQNVVDH-----MAN 288  
DB 558 TISPVPTKLVTQSPSERQIEQMPNSFRELQDSKFKPAQAPNPKPERVERGSAEQSMSE 617  
QY 289 HLGVPNRIILLGSESLSTPTATPSTLK-----LGVADIIDCVILA-----SS 331  
DB 618 PLIS-----RVARG-SPIAPKPRSPLOAPLLETATPTPIDTAETAIERSAEFSS 672  
QY 332 SEATETSQEL 341  
DB 673 SHSMDPSNSL 682

## RESULT 8

WMBEDE  
65K early nonstructural protein - human cytomegalovirus (strain AD169)  
N:Alternate names: UL84 protein  
C:Species: human cytomegalovirus, human herpesvirus 5  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 16-Jul-1999  
A:Accession: S09848  
R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; M.; Barrell, B.G.  
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
A:Reference number: S09749; MUID:90269039; PMID:2161319

## A:Accession: S09848

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-586 &lt;CHE&gt;

A:Cross-references: EMBL:X17403; NID:G59591; PIDN:CAA35358.1; PID:g1780862

A:Note: this sequence was submitted to the EMBL Data Library, December 1989

C:Superfamily: human cytomegalovirus 65K early nonstructural protein

C:Keywords: leucine zipper; nonstructural protein

F:114-135/Region: leucine zipper motif

F:324-372/Region: leucine zipper motif

Query Match 6.3%; Score 132; DB 1; Length 586;

Best Local Similarity 23.6%; Pred. No. 1.1;

Matches 104; Conservative 62; Mismatches 145; Indels 130; Gaps 25;

QY 6 RGRGPRSR---GGGARRARGRCRRAROS-----PARLPDPTVLVDL--VSDSDEEV 55

DB 11 RARRPARRGGGGVGSNSRHS-GKRRQRALSAPPLTFLATTTTMMGVASTDDDS 69

QY 56 LEVADPVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLDLPGEAPVVP 115

DB 70 LLKTPDEL-----DKYSGSPQTL-----TLTKHDIRQPR 101

QY 116 VYSGK---VOSSINLIPDN-----SSLKLCPEPEDEADLTNSGSSPSEDALFS 163

DB 102 VHRGTVHLQLHLDRPELRDPFOILLSTPLQL--GEANDESQ--TAPATLQEEETAAS 157

QY 164 GSPWRKLRKKKEKKMEEPDODISPLQPSRNKSRKHTKALQKLVNKRLODLR 223

DB 158 HPEKKK-----EKQKKEE--DED-----DRNDRER-----GILCVSNESDVR 197

QY 224 SCLS--PKHQSPALQSTDDEVVLVEGVPVLPQSSRLFTKIRCADLVRLPVRMSEPLQN 281

DB 198 PAFSLFPAPRGCHILRSVIDQ-QLTMAIVRLSLNLFALRI-ITPLKRLPLR-----RK 250

QY 282 VVDHMANHGVSPNRIILLFGSELSPTATPSTLKLGVADIIDCVVLASSSEATETSQEL 341

DB 251 AAHTALH-----DCLALHLPETFEPLDINNVTENAAASVAD---TAESTDA-DLTPTL 301

QY 342 RLRVQGEKKHQMLEISLSPDSLKVLMSHYEAMGLSG-----HKLSD----- 393

DB 302 TVRVR-----HALCWHRVG-GISGRGLTSRISARLSETTAKTLGP 342

QY 384 FFFDGTGLSGKELPADLGLS 404

DB 343 SVFGRLELDPNESPPDLTSS 363

## RESULT 9

T09648  
nucleolin homolog nuM1 - alfalfa  
N:Alternate names: nuM1 protein  
C:Species: Medicago sativa (alfalfa)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
A:Accession: T09648  
R:Boegre, L.; Jonak, C.; Mink, M.; Meskiene, I.; Traas, J.; Ha, D.T.C.; Swoboda, I.; P1  
Plant Cell 8, 417-428, 1996  
A:Title: Developmental and cell cycle regulation of Alfalfa nucM1 a plant homolog of t  
A:Reference number: Z16796; MUID:96361876; PMID:8721748  
A:Accession: T09648  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-635 <BOE>  
A:Cross-references: EMBL:X88845; NID:g1279562; PIDN:CAA61298.1; PID:g1279563  
C:Genetics:  
A:Gene: nuM1  
C:Superfamily: nucleolin; ribonucleoprotein repeat homology

Query Match 6.1%; Score 128.5; DB 2; Length 635;

Best Local Similarity 20.8%; Pred. No. 1.9;

Matches 92; Conservative 75; Mismatches 183; Indels 93; Gaps 20;

QY 24 AGRGCRPRQSPA-----RLIPDPTVLVDLSD---SDEVLEVADPVEVPVARLPAPAK 74



```

      80 AKAPAP-SKPTPAKGNVKKAOPETTSESDDISSDEEVKPAKAVPSKNGSAPVK 138
      75 -----PEODSDSGAAGPAGAPRTLVRRRRRLDPGEAPVVPVSGVQSSLNILIP 129
      Db 139 KODESEEDSDSSSDKPKPAKAVPSKN-----GSAFAKKDDSEEDSDSDSE 191
      Qy 130 DNSLLKLCPSF-----PEDEADLTNGSGSPSEDDALPS-----GSPWRKKRKCE 176
      Db 192 DEKPAKAVPSKNGSVAKKDDSSDESDSEEDKPKPAKASKNVSAPTKKAASSSDE 251
      Qy 177 KREKMEEPPODDISPLQPPSS-RKSRKHTALOKLREVNRLQDLRSCSPKQHQSPA 235
      Db 232 BSDESSDE--DEDAKPVSKPAVAKSKDSSDSDDEDDSSSDKPKVASKKEVSESE 309
      Qy 236 LOSTDEVLVEGPVLPOSSRLFTLKIRCADVLRLPVMSBPLONVVD-----HMANIL 290
      Db 310 SSSSDDDHKM---NIDKSSDSEDESEDEDEPLKTPQKKIRQDVE-MYDAGSKGKAPRT 365
      Qy 291 GVSPPRILLFGESLSPTAPSTLKL-----GVADIIDCVVLASSSE----- 333
      Db 366 PATPN-----ETSGSKTLFVGNLSFDVKRSDIENFGCGVEVVD-VRLASDGGVFKG 417
      Qy 334 -----ATETSGELRLRVQGE-KHQMLEISL-----SPSPKLKLMSHYEAMGLSG 379
      Db 418 FGVHFAFAEAQAQALENMGQELHRLALDLARERGAFTPNN-----NSNYSAGSGRG 472
      Qy 380 HKLSFFPDG--TKLSGKELPADL 400
      Db 473 OSQTVFVRGFDKNGLEDERAKL 495

```

# RESULT 10 B42680 nucleolus-cytoplasm shuttle phosphoprotein - rat

```

N:Alternate names: Noppi140 protein B
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 17-Nov-2000
C:Accession: B42680; A42680; S27889; S27890; S30510; S30511
R:Meier, U.T.; Blobel, G.
Cell 70, 127-138, 1992
A:Title: Noppi140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: A42680; MUID:92323542; PMID:1623516
A:Accession: B42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-734 <ME1>
A:Cross-references: GB:M94288
A:Experimental source: clone pTW6
A:Accession: A42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 32-180, 'O', 181-734 <ME3>
A:Cross-references: GB:M94287; NID:9205749; PIDN:AAA41718.1; PID:9205750
R:Meier, U.; Blobel, G.
submitted to the EMBL Data Library, May 1992
A:Description: Noppi140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: S27889
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-734 <ME2>
A:Cross-references: EMBL:M94288; NID:9205751; PIDN:AAA41719.1; PID:9205752
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-180, 'O', 181-734 <ME1>
A:Cross-references: EMBL:M94287; NID:9205749; PIDN:AAA41718.1; PID:9205750
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

```

Query March 6.1%; Score 128.5; DB 2; Length 734;  
Best Local Similarity 23.2%; Pred. No. 2.3; Indels 95; Gaps 12;  
Matches 71; Conservative 35; Mismatches 105; Indels 95; Gaps 12;

```

      Qy 33 QSPARLIPDTVLVDLSNDER-----VLEVADPVEVPAARL---PAPAKPEODSDSDSE 84
      Db 347 OSPKAAAQOTOPADSSADSSSESDSSSEEEKTPAKTIVSKTPAKPAPVKKKAESSSDS 406
      Qy 85 GAAGBPAGAPRTLVRRRRRLDPGEAPVVPV-----YSGKVOSSLNL 127
      Db 407 DSDSDEAPAKPVATSPSLSKPAVTGPKPAKAVATPKQAGSGQKQSKAASSSE 466
      Qy 128 IPDNGS-----LTKLCSEPEDEA-----DLTNGSGSPSEDD----- 159
      Db 467 BESSSSSEBATKSVTTTKARVATAKAAPSLPAKQAPRAGGSDSSDESSSEEEKTPPK 526
      Qy 160 -----ALPSCSPVRKKLRK-----KCEKEKKMEEPPODDISPLQ--- 195
      Db 527 PPAKKKAAGAAVPKPTPVKKAASESSSSSSSESDSESEEEKK---PKSKATPKQAGKA 582
      Qy 196 ---PSSRN-KSRKHTALOKLREVNKRLODLRSCSPKQHQSPALOSTDEVLVEGPVL 251
      Db 583 NGVPASQNGKAGKESEEBEDTEONKKAAGTKPGSGKKRKN---ETADEA-----AT 632
      Qy 252 PQSSRL 257
      Db 633 PQSKRV 638

```

# RESULT 11 D87719 protein R119.4 (imported) - Caenorhabditis elegans

```

C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence revision 10-May-2001 #text change 10-May-2001
C:Accession: D87719
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99065613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Accession: D87719
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-742 <STO>
A:Cross-references: GB:chr_I; PIDN:AACT6432.1; PID:93133012; GSPDB:GN00019; CBSP:R119.4
C:Genetics:
A:Gene: R119.4
A:Map position: 1

```

Query Match 6.1%; Score 128.5; DB 2; Length 742;  
Best Local Similarity 23.0%; Pred. No. 2.3;  
Matches 78; Conservative 33; Mismatches 111; Indels 117; Gaps 14;

```

      Qy 7 GRGPRS-----RGGRGARARGRCPRARQSPARLIPDTVLVDLSDS---D 52
      Db 170 GAAPSTRGCFEKKPYVGGRG-----GARGYSRAVAPSGALEPDATADDEMQTKVD 224
      Qy 53 BEVLEVADPVE-----VPVRLPAPAKPEODSDSGAAGPAGAPRTLVRRRRRL 106
      Db 225 TTIVTEVQPVESVATVPTTSAPAPL-----SFAVAAAHAHKELARKKQAQNP 274
      Qy 107 DPGAPVVPVYSGKVOSSLNLIPDNGSLKLCPSPEDEADLTNGSGSPSEDDALPSCSP 166
      Db 275 QPAPAP-----RRSLSPQPP-----LPSVAP 295
      Qy 167 WRKK-----LRKKCEKEKKMEEPPODDISPLQPPSSRNKRGHTALOKLREVNRL 219
      Db 296 VKEEPAAPVFPPEPTSTQPEKEDFPQNESSVLAEEQTPNVSTHDEVQS----- 347
      Qy 220 QDLRSCSPKQHQSPALQ-STDEVLVEGPVLPOSSRLFTLKIRCADVLRLPVRMSEP 278
      Db 348 -----TPEPQAWTQKTDLDLIGLSEAPQLGSS-----PLPSAP 383
      Qy 279 LQNVVDHMANHLG-VSPNRIL-LRGESELSPT-ATPST 314
      Db 384 VQIIPDPGVFEVGTATPTNIDYSPGFVEAAPSPLPST 422

```



```
QY 101 RRRLLD---PGEARVPVYSGKVOSSLNLIIPDNSS---LLKLCPSPEDEADILTNGS 153
Db 483 PTKKPADPKPSPEDPKVPAPKMKPMPWEDDDDEPADFTVPKPGEDEDEPADDBE 542
QY 154 SPSEDDALPSSGPMKRIKRCCEKEKCEEPDODISLPDP----- 196
Db 543 EPEDP---PAEDPPEKKPKPKRKRKKKPPVEPEKEPTPEPVVPKAPKIAPKKPBE 599
QY 197 -----SSRNKSRKHTALOKLREVNKRLQDLRSCLSPKOHSPAL----- 236
Db 600 PIPMPKPKTIARERKERIPALRIYAKK-PRLEVIYIPVYPMW-QTAALITQEGMGAF 657
QY 237 ---OSTDDEVVLVEGPVL---QSSRLFTL-----KIRCRADLVRLPYRMSEPLQNVVDH 285
Db 658 GKSRANVEVNFGRPIVOGAVDSKTVIPLMNDKSKANRSGMTAFGA-PRIDQNVVDH 716
QY 286 MANHGVSPNRILLFGESELSPTATPSTLKGVADI 322
Db 717 --HKGKSQGIPLAKGTYHPHGEYGTIRQTDV 750
```

## RESULT 15

```
JM0057
Gravin - human
C/Species: Homo sapiens (man)
C/Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C/Accession: J00057
R/Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 119-1126, 1998
A/Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial
A/Reference number: J00057; MUID:98269042; PMID:9604001
A/Accession: J00057
A/Molecule type: mRNA
A/Residues: 1-1684 <SAT>
A/Cross-references: DDBJ:AB003476; NID:G2081606; PIDN:BA19927.1; PID:d1020716; PID:g208
C/Comment: This protein regulates cell growth.
F/433-439/Region: nuclear location signal
F/522-527/Region: nuclear location signal
F/591-596/Region: nuclear location signal
F/671-676/Region: nuclear location signal
```

Query Match 5.9%; Score 124.5; DB 2; Length 1684;  
Best Local Similarity 20.7%; Pred. No. 11;  
Matches 94; Conservative 70; Mismatches 171; Indels 119; Gaps 21;

```
QY 43 VLVVLSDSDEVLVAPVVPVAPRLPAPAKPEQDSDD----- 82
Db 32 VVHDITDGOETPPIIQIPSESNLELTQPTESQANDIGFKKVPKVFVKTKDK 91
QY 83 -----SEGAEGPAGA-----PRTLVRRRRLDPGEAPVPVYSGK-VQSS 124
Db 92 TEKPDVOLLTVKDKDEGACAGCDKPSLGAAGAAKSESPKOSTEKPEETLKREOSH 151
QY 125 LNLIP--DNSSLKLCPSPEPD--EADILTNGSSPSEDDALPSGSP-----WRK 169
Db 152 AEISPPAESQAVECEKEGEKEKPSKSAESPTSPTSSTGTFKKFTQGWAGWRK 211
QY 170 K--LRKKKEK-----EKMEEPDODISLPQPSRNKSRKHTALQKLRVNRKLDLR 223
Db 212 KTSFRKPEDEVAEKKKEQEPK---VTEEDGKAESAEXTLTASEQAHPQ-EPAE 265
QY 224 SCLSPK---OHSPALOSTDDEVVLVEGP---VLPSSRLFTLKIRCRADLVRLPYRMS 276
Db 266 SAHEBRLSAEYKVELPS-EEVYSSGSPSEKRPAPLATEVPDEKIEVHQEEVAEVAHS 324
QY 277 EPLQNVVDMANHLGVSPNRILLFGESELSPTATPSTLKGVADIIDCVVLASSSEATE 336
Db 325 TVEERTBEQ-----KTEVEETA-----GSVPABELVEMDAEPQEA 360
QY 337 TSQEL-RLR---VQCKEKGMLKELSLSPDSPL-----KVLMSHYEAMGLSGHKL 382
Db 361 PAKELVTKKETCVSGEDPTQAD--LSPDEKVLKRPGEVSEVEMLSQERMKVQSGSPL 418
```

```
QY 383 SFFPDGT---KLSGKEPLADIG---LESGLIEV 410
Db 419 KKLFTSTGLKKLSGKKQKQKRGGBESGERTQV 452
```

Search completed: January 12, 2003, 10:28:03  
Job time : 28 secs

This Page Blank (uspc)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 10:23:09 ; Search time 14 Seconds  
(without alignments)  
1220.589 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099  
Sequence: 1 MAEPDLRGPRSRGCGARR.....GKELPADLGESGDLEIWMG 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135.5	6.5	831	1 NPH_RAT	P16884 rattus norv
2	134.5	6.4	587	1 UH84_HCMWT	P28839 human cytom
3	132	6.3	586	1 UH84_HCMVA	P16727 human cytom
4	128.5	6.1	704	1 NP14_RAT	P41777 rattus norv
5	128	6.1	303	1 VG08_BP22	P26748 bacterioph
6	127	6.1	1365	1 SU22_DROME	P25172 drosophila
7	122.5	5.8	1781	1 AKAC_HUMAN	Q02952 homo sapien
8	121	5.8	1109	1 TCF8_RAT	Q62947 rattus norv
9	119	5.7	1603	1 PSC_DROME	Q35820 drosophila
10	118.5	5.6	718	1 TRF2_CHICK	Q95653 gallus gall
11	116.5	5.6	1039	1 MSU1_DROME	P50535 drosophila
12	116.5	5.6	1061	1 LUS_DROME	P34739 drosophila
13	116	5.5	1043	1 TCF8_MESAU	Q60542 mesocricet
14	115.5	5.5	872	1 SCDS_YEAST	P34758 saccharomyc
15	114.5	5.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
16	113.5	5.4	1280	1 DYNA_RAT	P28023 rattus norv
17	112.5	5.4	95	1 SM32_HUMAN	P18858 homo sapien
18	112	5.3	919	1 DNLI_HUMAN	Q01188 homo sapien
19	112	5.3	943	1 CENC_HUMAN	Q01188 homo sapien
20	112	5.3	1004	1 PRPX_RAT	Q64545 rattus norv
21	111.5	5.3	1170	1 XPG_MOUSE	P35689 mus musculu
22	111	5.3	741	1 BSG2_DROME	P11929 drosophila
23	111	5.3	897	1 EP15_MOUSE	P42567 mus musculu
24	111	5.3	1163	1 RTN4_RAT	Q91111 rattus norv
25	111	5.3	1281	1 DYNA_MOUSE	Q08788 mus musculu
26	110.5	5.3	854	1 CLPB_THERH	Q94633 thermus the
27	110.5	5.3	919	1 SYNP_HUMAN	Q94618 h nuclear r
28	110.5	5.3	2517	1 NCR2_HUMAN	Q94618 h nuclear r
29	110	5.2	542	1 TULI_HUMAN	Q00294 homo sapien
30	110	5.2	612	1 ARRS_MAIZE	P13027 zea mays (m
31	110	5.2	883	1 E74B_DROME	P11536 drosophila
32	110	5.2	1206	1 FM14_MOUSE	Q05859 mus musculu
33	110	5.2	2150	1 SDC3_CAEEL	P34706 caenorhabdi

34	109.5	5.2	536	1 GAG_MLVCB	P27460 cas-br-e mu
35	109.5	5.2	972	1 ORC4_SCHPO	Q97794 schizosach
36	109	5.2	974	1 YMB4_CAEEL	Q03601 caenorhabdi
37	109	5.2	1089	1 Y553_HUMAN	Q94613 homo sapien
38	109	5.2	3312	1 CLR3_HUMAN	Q94613 homo sapien
39	108.5	5.2	487	1 ATP2_CHICK	Q93602 gallus gall
40	108.5	5.2	505	1 TUB_MOUSE	P50586 mus musculu
41	108.5	5.2	1117	1 TCF8_MOUSE	Q64318 mus musculu
42	108	5.1	487	1 ATP2_HUMAN	P15336 homo sapien
43	108	5.1	487	1 ATP2_MOUSE	P16951 mus musculu
44	108	5.1	673	1 FXR2_HUMAN	P51116 homo sapien
45	108	5.1	1402	1 IF4G_RABIT	P41110 oryctolagus

## ALIGNMENTS

RESULT 1  
ID NPH\_RAT STANDARD; PRT; 831 AA.  
AC P16884; Q63368;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neurofilament triplet H protein (200 kDa neurofilament protein)  
DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).  
GN NEFH OR NPH.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
ON [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=89065087; PubMed=3143606;  
RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;  
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).  
RT Identification of putative phosphorylation sites.";  
RL FEBS Lett. 241:213-218(1988).  
RN [2]  
RP SEQUENCE OF 37-831 FROM N.A.  
RX MEDLINE=89309090; PubMed=2457365;  
RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nusbaum J.L.,  
RA Jolles P.;  
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and  
RT in situ detection.";  
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).  
RN [3]  
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.  
RX MEDLINE=87080760; PubMed=2878828;  
RA Robinson P.A., Wion D., Anderson B.H.;  
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide  
RT (NF-H).";  
RL FEBS Lett. 209:203-205(1986).  
RN [4]  
RP SEQUENCE OF 318-831 FROM N.A.  
RX MEDLINE=89184647; PubMed=2928342;  
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldsaber D.,  
RA Smolowitz W., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;  
RT "Cloning of a cDNA encoding the rat high molecular weight  
RT neurofilament peptide (NF-H): developmental and tissue expression in  
RT the rat, and mapping of its human homologue to chromosomes 1 and  
RT 22.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).  
RN [5]  
RP FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
RP AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
RP NF-H HAS AN IMPORTANT FUNCTION IN NATURE AXONS THAT IS NOT  
RP OBSERVED BY THE TWO SMALLER NF PROTEINS.  
RP -!- PM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NPH IS  
RP PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
RP THOUGHT THAT PHOSPHORYLATION OF NPH RESULTS IN THE FORMATION OF  
RP INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
RP OF AXONAL CALIBER.



```

Db 257 ALHDCMALHL-----BELTFESTLIDINNNTENASVADAESTDADLT 299
Qy 339 QELRLRVOGKEKQMLLEISLSPDPLKV-----LMSHYEAMGLSGHKLSPFFDGTCLSG 393
Db 300 PLTVRVHVAOCMHVEGIGSGPRGLTSRISARLSETTAKTIGPS-----VFGRLLELDP 353
Qy 394 KELPADLGLES 404
Db 354 NESPDLTLLSS 364

RESULT 3
UL84 HCMVA STANDARD; PRT; 586 AA.
ID UL84 HCMVA
AC P16727;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE 65 kDa early nonstructural protein (UL84 protein).
GN UL84.
OS Human cytomegalovirus (strain AD169).
OC Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchison S.C., Ili, Kouzarides T., Martignetti J.A.,
RA Predlie E., Satchwell S.C., Tomlinson P., Weston K.M., Bartell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X17403; CAAS3558.1; -.
DR PIR; S09848; WMBEDE.
KM Nonstructural protein.
FT DOMAIN 9 19 ARG-RICH (BASIC).
FT DOMAIN 162 170 LYS-RICH (BASIC).
FT DOMAIN 171 182 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 586 AA; 65428 MW; 54AB912D6077223F CRC64;

Query March 6.3% Score 132; DB 1; Length 586;
Best Local Similarity 23.6%; Pred. No. 0.47; Indels 130; Gaps 25;
Matches 104; Conservative 62; Mismatches 145;

Qy 6 RGRGRSR-----GGGARARAGARCRPARQS-----PARLIDTVLVDL--VSDSDEEV 55
Db 11 RARPRARAGGGGAGVSSSRHS-GKCRORRALAPLFLATTTTMMGVASTDDDS 69
Qy 56 LEVADPVVPAARLPAPAKPEODSDSDEGAEGPAGPRTLVRRRRLDPCGAPVVP 115
Db 70 LLKTPDEL-----DKYSGSPOTIL-----TLTDKHIROPR 101
Qy 116 VYSGK--VQSSSLNIPDN-----SSLKLCPSPEPEADLTNGSSPEDDALPS 163
Db 102 VHRGTYHILQHLDRPELDPFOILSTPOL--GEANDESO--TAPATLOEETAAAS 157
Qy 164 GSPRRKRLRKKCEKEKKMEFPDODISPLPOPSRRNKRKHTALQKLRVNRKRLQDLR 223
Db 158 HEPEKKK-----EKQEKKE--DED-----DRNDRRER--GILCVSNSDSQVR 197
Qy 224 SCLG--PKQHOSPALQSDDEVDVVEGVLPQSSRLFLTKRCRADVRLPYRMSEPIQN 281
Db 198 PAFSLFPAFGCHILRSVIDO-QLTRMAIVRLSLNLFALRI-ITPLKRLPUPR-----RK 250

```

```

Qy 282 VVDHANHLGVSPNRILLFESSELSPATPSTLKGVAIDIIDCVLLASSSEATSTQSEL 341
Db 251 AAHHTALH-----DCLALHLPELTFEPLDINNNTENASVAD---TAESTD A-DLTPTL 301
Qy 342 RLRVQKEKQMLLEISLSPDPLKVMHYEAMGLSG-----HKL----- 383
Db 302 TVRV-----NALCWHRVEG-GISGPRGLTSRISARLSETTAKTIGP 342
Qy 384 FFPDGTKLSGKELPADLGLES 404
Db 343 SVFGRLLEDPNESPDLTLLSS 363

RESULT 4
NP14 RAT STANDARD; PRT; 704 AA.
ID NP14 RAT
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nop140) (Nucleolar and coiled-body
DE phosphoprotein 1).
GN NOLC1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RX MEDLINE=92323542; PubMed=1623516;
RA Meier U.T., Blobel G.;
RT "Nop140 shuttles on tracks between nucleolus and cytoplasm."
RL Cell 70:127-138(1992).
CC [2]
CC INTERACTION WITH NOP5 AND FIBRILLARIN.
CC MEDLINE=20143579; PubMed=10679015;
CC Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
CC "Conserved composition of mammalian box H/ACA and box C/D small
CC nucleolar ribonucleoprotein particles and their interaction with the
CC common factor Nop140."
CC Mol. Biol. Cell 11:567-577(2000).
CC -1- FUNCTION: RELATED TO NUCLEOGENESIS. MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY).
CC -1- SUBUNIT: Interacts with DKC1/Nap57, NOP5/Nap65 and fibrillarin.
CC -1- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN
CC NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE
CC FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOLAR PLASM TO
CC A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.
CC -1- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOP140 IS ONE OF THE
CC MOST PHOSPHORYLATED PROTEINS IN THE CELL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M94287; AAA41718.1; -.
DR EMBL; M94288; AAA41719.1; -.
KM Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.
FT DOMAIN 84 570 11 X 12 AA APPROXIMATE REPEATS OF AN
FT REPEAT 84 95 ACIDIC SERINE CLUSTER.
FT REPEAT 127 138 ACIDIC SERINE CLUSTER 2.

```







```
FT DOMAIN 752 782 AKAP 2.
FT DOMAIN 797 827 AKAP 3.
FT DOMAIN 98 101 POLY-GLU.
FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).
FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).
FT VARSPPLIC 1 98 MISSING (IN ISOFORM 2).
FT VARSPPLIC 99 106 EREVIVTE -> MLGTITIT (IN ISOFORM 2).
FT CONFLICT 117 117 E -> K (IN REF. 2).
FT CONFLICT 142 144 NRN -> TPEI (IN REF. 2 AND 3).
FT CONFLICT 215 215 Q -> K (IN REF. 2).
FT CONFLICT 448 448 G -> E (IN REF. 2 AND 3).
FT CONFLICT 694 694 R -> G (IN REF. 2 AND 3).
FT CONFLICT 867 867 G -> S (IN REF. 2 AND 3).
FT CONFLICT 986 986 S -> A (IN REF. 2 AND 3).
FT CONFLICT 1530 1530 E -> EE (IN REF. 3 AND 4).
FT CONFLICT 1581 1581 V -> M (IN REF. 4).
FT CONFLICT 1601 1601 Q -> L (IN REF. 2).
SQ SEQUENCE 1781 AA; 191439 MW; BA813937379FACOF CRC64;

Query Match 5.8%; Score 122.5; DB 1; Length 1781;
Best Local Similarity 22.0%; Pred. No. 6.9;
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;

QY 40 PDTV-LVLDVSDDEVEVADPVEVPVARLPAPAK---PEQSDSDSEGAAGPAGAPR 95
DB 192 PDTVQLLVKKDEGEAGAGDHPDPSLGAAGEAASKESEPKQTEKPE-----240
QY 96 TLVRRRRRLDPGEAPVVPVYSGKQSSLNLIIPDNSSLLKLCPSPEP---EADLTNSGS 153
DB 241 TLKREOSHAEISP-----PAESGQA-----VEECKEESGEKQEKESKSAE 281
QY 154 SPEDDALPGSP-----WRKK--LRKKCKPE-----EKKWEEPPDQDISPLPOP 196
DB 282 SPTSPVTSGETGTFKKFFTOGWAGRRKTSFRKPKEDVEASEKKKEQPEK-----VDT 336
QY 197 SSRNKRKHTALQKLEVNKRLODLRSLSPK---QHSPALQSTDDDEVVLVEGP-----249
DB 337 EEDGKAFAVESEKLTASEQAHPQ-EPAESAEHPERLSAEYKVELPS-EEQVSGSQSPSEK 394
QY 250 VLPQSRLFTLTKRCRADLRLPVRMSEPLQNVVDHMANHLGVSPNRIILLFGESELSPT 309
DB 395 PAPLATEVDFEKLVEHQEEVVAEVHVSTVEETEEO-----KTEVEET 437
QY 310 A--TPSTLKLGVADIIDCVVLASSSEATETSQBLRLR---VOGKEKHQMLEISLSPDPL 364
DB 438 AGSVPAEELVGM-----AEPQAEAPAKELVKLKETCVSGEDPTQGAD--LSPDEKV 487
QY 365 -----KVLMSHYEEAMGLSHGKLSFFDGT---KLSGKELPADLG---LESGDLI 408
DB 488 LSKPPEGVSEVEMLSQERMKVQSGPLKLTFTSTGLKLSGKKQKRGKGDESGEHT 547
QY 409 EV 410
DB 548 QV 549

RESULT 8
TCF8_RAT STANDARD; PRT; 1109 AA.
ID TCF8_RAT
AC Q62947; Q62948;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor 8 (zinc finger homeodomain enhancer-binding protein) (Zfhep).
DE protein) (Zfhep).
GN TCF8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 6-1109 FROM N.A.; AND ALTERNATIVE SPLICING.
RX MEDLINE=96365389; PubMed=8769566;
```

```
RA Cabanillas A.M., Darling D.S.;
RT "Alternative splicing gives rise to two isoforms of Zfhep, a zinc
RL finger/homeodomain protein that binds T3-response elements.";
RL DNA Cell Biol. 15:643-651(1996).
CC -!- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC SEQUENCES IN THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/ZFHEP-1 (SHOWN HERE) AND
CC 2/ZFHEP-2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMEODOMAIN PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U51583; AAB17130.1; -.
CC EMBL; U51584; AAB17131.1; -.
CC InterPro; IPR0001356; Homeobox.
CC InterPro; IPR0000822; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 7.
CC ProDom; PD000003; Znf_C2H2; 2.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC SMART; SM00355; Znf_C2H2; 7.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
CC Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW Homeobox; Repressor; Activator; Metal-binding; Repeat;
KW Alternative splicing.
FT ZN_FING 150 173 C2H2-TYPE.
FT ZN_FING 180 202 C2H2-TYPE.
FT ZN_FING 220 242 C2H2-TYPE.
FT ZN_FING 248 272 C2H2-TYPE (ATYPICAL).
FT DNA_BIND 559 618 HOMEBOX-LIKE.
FT ZN_FING 881 903 C2H2-TYPE.
FT ZN_FING 909 931 C2H2-TYPE.
FT ZN_FING 937 958 C2H2-TYPE (ATYPICAL).
FT DOMAIN 968 1109 GLU-RICH (ACIDIC).
FT VARSPPLIC 1 198 MISSING (IN ISOFORM 2).
SQ SEQUENCE 1109 AA; 121626 MW; BEPE291C8795DDA6 CRC64;

Query Match 5.8%; Score 121; DB 1; Length 1109;
Best Local Similarity 21.4%; Pred. No. 4.7;
Matches 87; Conservative 57; Mismatches 152; Indels 110; Gaps 19;

QY 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRRLDPGEAPVVPVYS-G 119
DB 535 PAQPPP---PAPATEKPESASAGND-----LSPQPLKNLLSL 574
QY 120 KVQSSNLIPDNSSLLKLCPS--PEDEA----DLTNSGSPSEDDALPSGSPWRKLRK 173
DB 575 KAYALNAQPSSTEELTKIADSVNLPLDVVKWFKWQAGQIPQSLEPPSPGSGNIPA 634
QY 174 KEKEEKWE-EFPQD-----ISP-LPQSSRNKRKHTALQKLEVNKRLO 220
DB 635 KTEEQPQVDPGNEPQBDSTRGOSPLKMTSSPVLPGVSAINGSRCTSSPSPLNLSARNP 694
QY 221 DLRSCLSPKHQSPALOSTDDEVVLVEGPVLPQS--SRLF-----TLKIRCADL 268
DB 695 QGVSCVSEGTQEPQVEPDLPLSLPKQGGELLERSTVSSVQNSVYSVQEEPLNLSAK-- 752
QY 269 VRLP-----VRMSEPLQNVVDHMANHLGVSPNRIILLFGESELSPTAT---PSTKLGLVA 320
DB 753 -KEPKDSCVTDSEPVVNVVPPSANPINIA-----IPTVTAQLPTIVAIDQ 798
QY 321 DIIDCVVLASSEATETSQELRLRVQGEKHQML-----EISLSP---DSPLKVLN- 368
```

```

Db 799 NSVPC-----LRALANKQTLLIPQVATYSATVSPAMQEPVAVIOP 841
Qy 369 -SHYEFAMGLSGHKLSTF---FFDGTGLSGKEPLADLSEGLIE 409
Db 842 NQNDQRODTSESGVSVEDQNDSDCTPRKKTKRKAMGNVACDLCD 887

RESULT 9
PSC_DROME
ID PSC_DROME STANDARD: PRT: 1603 AA.
AC P35820.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Posterior sex combs protein.
GN PSC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018190; PubMed=1833647;
RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
RL encode proteins with homology to the murine bml-1 oncogene.";
RL Nature 353:351-353(1991).
CC -1- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR
CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59375; CAA41965.1; .
DR PIR; S17983; S17983.
DR FlyBase; FBgn0005624; Psc.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; ZF-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT DOMAIN 47 53 POLY-THR.
FT DOMAIN 83 88 POLY-THR.
FT DOMAIN 91 98 POLY-THR.
FT DOMAIN 145 152 POLY-THR.
FT DOMAIN 184 202 POLY-SER.
FT ZN_FING 265 304 RING-TYPE.
FT DOMAIN 642 651 POLY-SER.
FT DOMAIN 1066 1069 POLY-GLY.
FT DOMAIN 1185 1189 POLY-PRO.
FT DOMAIN 1214 1217 POLY-PRO.
FT DOMAIN 1391 1396 POLY-PRO.
FT DOMAIN 1458 1461 POLY-ALA.
FT DOMAIN 1517 1520 POLY-GLY.
SQ SEQUENCE 1603 AA; 169999 MW; 77024F4097736473 CRC64;

Query Match 5.7%; Score 119; DB 1; Length 1603;
Best Local Similarity 18.9%; Pred. No. 9.8;
Matches 68; Conservative 54; Mismatches 123; Indels 114; Gaps 13;
Qy 45 VDLVSDSBEVLEVADPVEVPAVLPAKPKPEQSDSDS---DSEGAEGPAGAPRTLVRRR 101

```

```

Db 598 IDLSKQNSVTIIDMSDPERREIVK---PLKPEKESRKKKDXGSPKSSSSSSSSGCR 654
Qy 102 RRLRLDGPAPVAVPYSGVOSS---LNLIP-----DNSLLKTCPEEP----- 142
Db 655 KKKSPSLTPVPLTRITERINSPSGVSTLSPRVYTGATSEDPKSEPLKFKPKVKE 714
Qy 143 EDEADLTNNGSSPSEDDALPSPGSPWRKKLRKKCEKKEEFPQDQISPLPQPSRNKS 202
Db 715 SPERTLNRRATPPSPVQGSASP-----KSKGNLNDQSLKKPSCMPKSIASKRS 769
Qy 203 RKTALOKLRVKNRKLQDLRSCLSPKQHSPPALOSTDEVVL-----VEG 248
Db 770 KEPVAVSK-----KQKLSPLPTVPFKIRLPVTNGSSGTASPKIEK 812
Qy 249 PVLQSSR---LFTLKIRCRADVLRLP-----VRMSEP-----LQNVVDHMA- 287
Db 813 PUMPPPAKPMPLAPRKLQPSAQFAPPSPFIHHAGVMSAPGNRTPIAKRYQITLPKASR 872
Qy 288 -NHLGVSPNRILLP-----GESELPTATPST 314
Db 873 PVFPANIPNDVRLKADGTEIKSIGGGSVENNSNSAQKPHLYGPKGTVMGPALPAT 931

RESULT 10
TRF2_CHICK
ID TRF2_CHICK STANDARD: PRT: 718 AA.
AC Q9PU53.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Telomeric repeat binding factor 2 (TTAGCG repeat binding factor 2)
DE (Telomeric DNA binding protein).
GN TRF2 OR TRF2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.; AND VARIANT VAL-438.
RC STRAIN=White Leghorn; TISSUE=Embryo;
RX MEDLINE=20035814; PubMed=10571037;
RA Konrad J.P., Mills W., Easty D.J., Farr C.J.;
RT "Cloning and characterisation of the chicken gene encoding the
RL telomeric protein TRF2.";
RL Gene 239:81-90(1999).
CC -1- FUNCTION: Binds the telomeric double-stranded TTAGGG repeat.
CC Protects against end-to-end fusion of chromosomes and plays a role
CC in successful progression through the cell division cycle (by
CC similarity).
CC -1- SUBUNIT: Homodimer. Binds to RAP1 (by similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear. Colocalizes with telomeric DNA in
CC interphase cells and is located at chromosome ends during
CC metaphase (by similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in embryo.
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC -1- CAUTION: The sequence was modified by adding the 2 missing N-
CC terminal residues based on the sequence of the genomic clone.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ133783; CAB56220.1; .
DR HSSP; P54774; 1BA5.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA_binding; 1.
DR SMART; SM00395; SANT; 1.

```

```

DR PROSITE; PS50090; MYB_3; 1.
KW Cell cycle; Nuclear protein; Chromosomal protein; Telomere;
KW DNA-binding; Repeat; Polymorphism.
FT DOMAIN 24 220 DIMERIZATION (BY SIMILARITY).
FT DOMAIN 257 451 15 X 13 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 257 269 1.
FT REPEAT 270 282 2.
FT REPEAT 283 295 3.
FT REPEAT 296 308 4.
FT REPEAT 309 321 5.
FT REPEAT 322 334 6.
FT REPEAT 335 347 7.
FT REPEAT 348 360 8.
FT REPEAT 361 373 9.
FT REPEAT 374 386 10.
FT REPEAT 387 399 11.
FT REPEAT 400 412 12.
FT REPEAT 413 425 13.
FT REPEAT 426 438 14.
FT REPEAT 439 451 15.
FT DOMAIN 545 550 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DNA_BIND 660 713 MYB.
FT VARIANT 438 438 A -> V.
SQ SEQUENCE 718 AA; 80163 MW; DD6LF253A6912551 CRC64;

Query Match 5.6%; Score 118.5; DB 1; Length 718;
Best Local Similarity 22.9%; Pred. No. 3.9;
Matches 93; Conservative 49; Mismatches 136; Indels 129; Gaps 21;

QY 3 EPLRGGRSRGG-RCARRARG--RCPRARQSPARLPDVTVLVDVSDSDSEEVLEVA 59
DB 247 EPLR-RVKHSVGLRRAETAGGVAGAPSCPEMAKDP-----TGAPHHVGTVK 292
QY 60 DPVEVPVARLPAPAKEQSDSDSEGA-----AEGPA-----GAPR-TLVRERRR 103
DB 293 DAV-----RAPCPAESTDSQGTTRCAETARDVNGAPSPSEMTKOLLGAPKCTETARDVV 347
QY 104 RLDDGPEAPVVPVYS-GKQVSSLNL--IPDNSSLLKLCFSEPEDEADLTNSGSSPSDEDA 160
DB 348 RAPSPAESTKDPVGTGHAETARDVARAPSPAETTKNLGAPCACTVKNTRVAPSPAER 407
QY 161 LPSGSPWRKRLKRCBKEEKKMEFPDQDISPLPQSSRNKSRKTEALQKUREVNRKIQ 220
DB 408 -----RKDLVRAPKRAE--TARDVVRAPSPAER-----VKDTAGASE 442
QY 221 DLRSCLSPKQHOHPALQSTDDVEVLVEGVPVLPQSSRLFTLKIRCRADLVLRVPMSE--- 277
DB 443 PMKSASYPTASQ-PRIAAV-----KSSKVF-----VPVEISEQPA 477
QY 278 ---PLQNVVDHMANHLGV-----PNRILLFGSESEL-----SPTATPSTKLKG 318
DB 478 AAPV-----HAGVSSRDLETPFRVTYTYGIVLREAFKMLNSPDSDALFNKLD 528
QY 319 VADIICVVLASSEATSTQELRLRVQGEKQHMLEISLSPSPSLK 365
DB 529 ETDL-----PSPQQMSPSVSHRTKRRKEKNQSGSETLDSPEIPHK 568

RESULT 11
MSL1_DROME STANDARD; PRT; 1039 AA.
AC P50335;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Male-specific lethal-1 protein.
GN MSL-1.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCB1_taxid=7227;
RN [1]

```

```

RP SEQUENCE OF 85-1039 FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93314941; PubMed=8325488;
RA Palmer M.J., Megner V.A., Richman R., Manning J.E., Kuroda M.I.,
RA Lucchesi J.C.;
RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
RT encodes a novel protein that associates with the X chromosome in
RT males.";
RL Genetics 134:545-557(1993).
RN [2]
RP REVISIONS. SEQUENCE FROM N.A.
RX MEDLINE=95300219; PubMed=7781064;
RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
RA Kuroda M.I.;
RT "Expression of msl-2 causes assembly of dosage compensation
RT regulators on the X chromosomes and female lethality in Drosophila.";
RL Cell 81:867-877(1995).
CC -!- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
CC DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MLE, MSL-1 AND
CC MSL-3 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL
CC PROTEINS REQUIRES ALL THE OTHER MSLs FOR WILD-TYPE X-CHROMOSOME
CC BINDING.
CC -!- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS
CC OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES
CC AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL
CC SITES IN MALES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L42514; AAA98918.1; -.
KW FlyBase; FBgn0005617; msl-1.
KW Nuclear protein.
FT CONFLICT 188 193 PLPPAA -> HCHLLP (IN REF. 1).
FT CONFLICT 492 492 L -> S (IN REF. 1).
FT CONFLICT 670 670 I -> M (IN REF. 1).
SQ SEQUENCE 1039 AA; 117412 MW; 4759E9B5EF6E9F14 CRC64;

Query Match 5.6%; Score 116.5; DB 1; Length 1039;
Best Local Similarity 22.2%; Pred. No. 8.2;
Matches 82; Conservative 55; Mismatches 141; Indels 91; Gaps 19;

QY 46 DLVSDSDEE-----VLEVADPVEVPVARLPAPAKP--EQSDSDSDEGAERGAG-APR 95
DB 400 DYNARTEELMQNVYLLLEL-DPTEKTCAPSAHSTPHQKQSSQAEIRKEGNQRITE 458
QY 96 TLVRRRRRLDPGEAPVVP--VYSGKVQSSNLNLPDNNSSLLKLCFSEPEDEADLTNSGS 153
DB 459 KLLQLKPEPMVDALAPILPKWVAFKKDKHELVES-----PEVPKQQ----- 503
QY 154 SPSEDDALPSGSPWRKLRKKCKEKKMEFPDQDISPLPQSSRNKS-----RKHTEAL 209
DB 504 -PHQDAIIVDHNAINQL-----EVPKPLDKPKQDQKDEQDQDGLDVRVPEQ 550
QY 210 QKLRVNRKRLQDLRSCLSPKQHQSPALQSTDDVEVLVEGVPVLPQSSRLFTLKIRCRADIV 269
DB 551 EDVRKVQKQETLKRQPEDAPK-HLPKA-----VAPKVTKTSRSTLPKANTADI- 598
QY 270 RLPVRMSEPLQNVVDHMANHLG-----VSPNRILLFGSESELSP-----TATPSTKLK 317
DB 599 -----KDAPAKVI---ANHQSTKTQTPVKTQRLQVIRQYEMHPDMRTGSSAPSDIR- 649
QY 318 GVADIICVVLASSEATSTQELRLRVQGEKQHMLEISLSPSPSLK-----LMSHY 371
DB 650 -KQKVDPV-----STPETKTKSKSILV--NDKTKTTSQSPDQDIDVETVRRKLAHL 702

```

```

Oy 372 BEAMGLSGH 380
Db 703 KEELISQSH 711

RESULT 12
IDS_DROME STANDARD; PRT; 1061 AA.
AC P34739;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable helicase lodestar.
GN Lds.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCB1_TaxID=7227;
RN (1)
RX MEDLINE=92009170; PubMed=1916263;
RA Girdham C.G., Glover D.M.;
RT "Chromosome tangling and breakage at anaphase result from mutations
RT in lodestar, a Drosophila gene encoding a putative nucleoside
RT triphosphate-binding protein.";
RL Genes Dev. 5:1786-1799(1991).
RN (2)
RP CONCEPTUAL TRANSLATION.
RX MEDLINE=93181281; PubMed=8382805;
RA Bork P., Koonin E.V.;
RT "An expanding family of helicases within the 'DEAD/H' superfamily.";
RL Nucleic Acids Res. 21:751-752(1993).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO
CC BE INTRODUCED IN POSITION 946 TO PRODUCE THIS ORF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62629; CAA44496.1; ALT_FRAME.
DR PIR; A40580; A40580.
DR Flybase; FBgn002542; lds.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00176; SNF2_N; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC_C; 1.
KW Nuclear protein; Helicase; ATP-binding.
FT NP_BIND 465 472 ATP (POTENTIAL).
FT SITE 603 606 DEAD BOX.
FT SITE 1061 1061 DEAH BOX.
FT SQ SEQUENCE 1061 AA; 118189 MW; E82AA64254342B80 CRC64;

Query Match 5.6%; Score 116.5; DB 1; Length 1061;
Best Local Similarity 22.2%; Pred. No. 8.4;
Matches 92; Conservative 50; Mismatches 144; Indels 129; Gaps 20;
Oy 29 PARASPARLIDTVLVDSDESEVADPEVVARLP-APAKPEQSDSDSE--- 84
Db 94 PSARTTKRKPL-----GIPSDSEDEDELQORALSPSRMSITGVPRQDLDDDBSEIERY 147
Oy 85 --GAEGPGAGRTLVRRRRRLDGEAPVVPVY---SGKVOSSLN-----LIP 129
Db 148 SDEVOGEGTEAP-----TAEA-VVPGYTTFQFAGNIQNDLHSTIGADSEVID 193

```

```

Oy 130 DNSSLKLCPEPEDEADLTNMGSSPSEDDALP-----SGSPWRKKLRK----- 173
Db 194 DSSGSDVLILSNKETPIRILSS-----TDDDATNNKEMSGPFERPSKLSPPSSAGASV 249
Oy 174 -KCEK--BEKKKEFPDDISPLPPSSRANKSRK-----HTALQKLRPVNRLDGL 222
Db 250 VRTSKNLSQPTTQAVLKKQKTPPAFRSRIRKSEDKVVSQVYVDEMKRLAEKRVQSDA 309
Oy 223 RSCLSPKHOSPALQSTDEVLVGVLPPOSSRLFTLKIRCRADLVRLPYVMSPELQNV 282
Db 310 EKLFFKVAHKLPDKKS-----QIMKRIDTTLARELAMDE--QMI 345
Oy 283 VDHMANHLGVSPNRLILFGESSELSPTATPSTLKGVADIIDCVVLASSSEATETQGLR 342
Db 346 SALRVQGSNVPAVRVY-----KPTLNPRAPS-----IDTLMDDELSEAVNEIKPVY 392
Oy 343 LRVQG-----KEKHQMLEISLSP-----DSP--LKV-LMSHVEAM 375
Db 393 TGAQGMATFNQKALLTLBSLXDLHVLSDLPQPEVLADPEGLKVLNMHQRHAL 447

RESULT 13
TCF8_MESAU STANDARD; PRT; 1043 AA.
ID TCF8_MESAU
AC Q60542;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor 8 (Zinc finger protein BZF).
GN TCF8 OR BZF.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCB1_TaxID=10036;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95021206; PubMed=7935395;
RA Franklin A.J., Jettion T.L., Shelton K.D., Magnuson M.A.;
RT "BZF, a novel serum-responsive zinc finger protein that inhibits gene
RT transcription.";
RL Mol. Cell. Biol. 14:6773-6788(1994).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC SEQUENCES IN THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMEDOMAIN PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13856; AAC37667.1; -.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 7.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2; 2.
DR PRODOM; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW Homeobox; Repressor; Activator; Metal-binding; Repeat.
FT ZN_FING 94 117 C2H2-TYPE.
FT FT ZN_FING 124 146 C2H2-TYPE.

```

[illegible]

DB 467 IK-----PSSPTETVNSGGLQPLKPTATGSANYLMKHQHSVNNPV 509

RESULT 15

TCOF HUMAN STANDARD: PRT: 1411 AA.

ID TCOF HUMAN 013428: 099408: 099860:

AC 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Treacle protein (Treacher collins syndrome protein).

GN TCOF1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCI\_TaxID=9606;

11

RP SEQUENCE FROM N.A.

RA MEDLINE=96154183; PubMed=8563749;

RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Lofus S.K.,

Bonner C.A., Kopivnikar K., Wasmuth J.J.;

"Positional cloning of a gene involved in the pathogenesis of

Treacher Collins syndrome";

Nat. Genet. 12:130-136(1996).

12

RP SEQUENCE FROM N.A.

RA MEDLINE=97228900; PubMed=9074926;

RA Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,

Dixon M.J.;

"Identification of the complete coding sequence and genomic

organization of the Treacher Collins syndrome gene.";

Genome Res. 7:223-234(1997).

13

RP SEQUENCE FROM N.A.

RA MEDLINE=97250498; PubMed=9096354;

RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,

Ashley J.A., Lovett M., Jabs E.W.;

"TCOF1 gene encodes a putative nuclear phosphoprotein that exhibits

mutations in Treacher Collins syndrome throughout its coding

region.";

Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).

14

RP VARIANTS L-439; V-810; V-113 AND G-1355, AND VARIANT TCS R-53.

RA MEDLINE=97195537; PubMed=9042910;

RA Edwards S.J., Gladwin A.J., Dixon M.J.;

"The mutational spectrum in Treacher Collins syndrome reveals a

predominance of mutations that create a premature-termination

codon.";

Am. J. Hum. Genet. 60:515-524(1997).

CC -1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS

SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF

CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000

LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALITY

SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS,

ARESTIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE

EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2)

LATERAL DOWARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH

COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND

ZYGOMATIC COMPLEX; (4) CLEFT PALATE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U84642; AAC51185.1; JOINED.

DR EMBL; U84643; AAC51185.1; JOINED.

DR EMBL; U84644; AAC51185.1; JOINED.

DR EMBL; U84645; AAC51185.1; JOINED.

DR EMBL; U84646; AAC51185.1; JOINED.

DR EMBL; U84647; AAC51185.1; JOINED.

DR EMBL; U84648; AAC51185.1; JOINED.

DR EMBL; U84649; AAC51185.1; JOINED.

DR EMBL; U84650; AAC51185.1; JOINED.

DR EMBL; U84651; AAC51185.1; JOINED.

DR EMBL; U84652; AAC51185.1; JOINED.

DR EMBL; U84653; AAC51185.1; JOINED.

DR EMBL; U84654; AAC51185.1; JOINED.

DR EMBL; U84655; AAC51185.1; JOINED.

DR EMBL; U84656; AAC51185.1; JOINED.

DR EMBL; U84657; AAC51185.1; JOINED.

DR EMBL; U84658; AAC51185.1; JOINED.

DR EMBL; U84659; AAC51185.1; JOINED.

DR EMBL; U84660; AAC51185.1; JOINED.

DR EMBL; U84661; AAC51185.1; JOINED.

DR EMBL; U84662; AAC51185.1; JOINED.

DR EMBL; U84663; AAC51185.1; JOINED.

DR EMBL; U79649; AAB40722.1; JOINED.

DR EMBL; U79645; AAB40722.1; JOINED.

DR EMBL; U79646; AAB40722.1; JOINED.

DR EMBL; U79647; AAB40722.1; JOINED.

DR EMBL; U79648; AAB40722.1; JOINED.

DR EMBL; U79649; AAB40722.1; JOINED.

DR EMBL; U79650; AAB40722.1; JOINED.

DR EMBL; U79651; AAB40722.1; JOINED.

DR EMBL; U79652; AAB40722.1; JOINED.

DR EMBL; U79653; AAB40722.1; JOINED.

DR EMBL; U79654; AAB40722.1; JOINED.

DR EMBL; U79655; AAB40722.1; JOINED.

DR EMBL; U79656; AAB40722.1; JOINED.

DR EMBL; U79657; AAB40722.1; JOINED.

DR EMBL; U79658; AAB40722.1; JOINED.

DR GeneW; HGNC:11654; TCOF1.

DR MIM; 606847; -.

DR MIM; 154500; -.

DR InterPro; IPR003993; treacle.

DR Pfam; PF03546; treacle; 3.

DR PRINTS; PRO1503; TREACLE.

KW Disease mutation; Polymorphism.

FT DOMAIN 89 97

FT DOMAIN 204 207

FT DOMAIN 616 619

FT DOMAIN 919 924

FT DOMAIN 1285 1289

FT DOMAIN 1375 1386

FT DOMAIN 1398 1405

FT VARIANT 53 53

FT VARIANT 439 439

FT VARIANT 810 810

FT VARIANT 1313 1313

FT VARIANT 1355 1355

FT CONFLICT 1312 1312

FT SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;

Query Match 5.8%; Score 114.5; DB 1; Length 1411;

Best Local Similarity 21.4%; Pred. No. 16;

Matches 84; Conservative 50; Mismatches 172; Indels 87; Gaps 16;

Qy 2 AEPPLRGGRPRRG-----GRCARARGARCRPAROSPARTLDIVLVDVSDSEVL 56

Db 715 ANPAAARAPSKGTTISAPGKVTAAQAQKORSPSKVPV-----NPQNSTV 762

Qy 57 EVADPEVVP-VARLPAPAK-----PEQDS-----DSDEGAEGPAGAPRTLVRRRRRL 105

DR EMBL; U84642; AAC51185.1; JOINED.

DR EMBL; U84643; AAC51185.1; JOINED.

DR EMBL; U84644; AAC51185.1; JOINED.

DR EMBL; U84645; AAC51185.1; JOINED.

DR EMBL; U84646; AAC51185.1; JOINED.

DR EMBL; U84647; AAC51185.1; JOINED.

DR EMBL; U84648; AAC51185.1; JOINED.

DR EMBL; U84649; AAC51185.1; JOINED.

DR EMBL; U84650; AAC51185.1; JOINED.

DR EMBL; U84651; AAC51185.1; JOINED.

DR EMBL; U84652; AAC51185.1; JOINED.

DR EMBL; U84653; AAC51185.1; JOINED.

DR EMBL; U84654; AAC51185.1; JOINED.

DR EMBL; U84655; AAC51185.1; JOINED.

DR EMBL; U84656; AAC51185.1; JOINED.

DR EMBL; U84657; AAC51185.1; JOINED.

DR EMBL; U84658; AAC51185.1; JOINED.

DR EMBL; U84659; AAC51185.1; JOINED.

DR EMBL; U84660; AAC51185.1; JOINED.

DR EMBL; U84661; AAC51185.1; JOINED.

DR EMBL; U84662; AAC51185.1; JOINED.

DR EMBL; U84663; AAC51185.1; JOINED.

DR EMBL; U79649; AAB40722.1; JOINED.

DR EMBL; U79645; AAB40722.1; JOINED.

DR EMBL; U79646; AAB40722.1; JOINED.

DR EMBL; U79647; AAB40722.1; JOINED.

DR EMBL; U79648; AAB40722.1; JOINED.

DR EMBL; U79649; AAB40722.1; JOINED.

DR EMBL; U79650; AAB40722.1; JOINED.

DR EMBL; U79651; AAB40722.1; JOINED.

DR EMBL; U79652; AAB40722.1; JOINED.

DR EMBL; U79653; AAB40722.1; JOINED.

DR EMBL; U79654; AAB40722.1; JOINED.

DR EMBL; U79655; AAB40722.1; JOINED.

DR EMBL; U79656; AAB40722.1; JOINED.

DR EMBL; U79657; AAB40722.1; JOINED.

DR EMBL; U79658; AAB40722.1; JOINED.

DR GeneW; HGNC:11654; TCOF1.

DR MIM; 606847; -.

DR MIM; 154500; -.

DR InterPro; IPR003993; treacle.

DR Pfam; PF03546; treacle; 3.

DR PRINTS; PRO1503; TREACLE.

KW Disease mutation; Polymorphism.

FT DOMAIN 89 97

FT DOMAIN 204 207

FT DOMAIN 616 619

FT DOMAIN 919 924

FT DOMAIN 1285 1289

FT DOMAIN 1375 1386

FT DOMAIN 1398 1405

FT VARIANT 53 53

FT VARIANT 439 439

FT VARIANT 810 810

FT VARIANT 1313 1313

FT VARIANT 1355 1355

FT CONFLICT 1312 1312

FT SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;

Query Match 5.8%; Score 114.5; DB 1; Length 1411;

Best Local Similarity 21.4%; Pred. No. 16;

Matches 84; Conservative 50; Mismatches 172; Indels 87; Gaps 16;

Qy 2 AEPPLRGGRPRRG-----GRCARARGARCRPAROSPARTLDIVLVDVSDSEVL 56

Db 715 ANPAAARAPSKGTTISAPGKVTAAQAQKORSPSKVPV-----NPQNSTV 762

Qy 57 EVADPEVVP-VARLPAPAK-----PEQDS-----DSDEGAEGPAGAPRTLVRRRRRL 105

Db 763 LARGPASVPSVGKAVATAAAQAQTGPBEDSGSSEESDEEAEETLAQAKPSGKTHOIRAA 822  
Qy 106 LDP-----GEAPVVPVYSGKVQSSNLNLPDNNSSLLKLCPSPEPEDEADLTNSG----- 152  
Db 823 LAPAKESPRKGAAPTFGKTGPSAAQAGKQDDSGSSSEESDSGGEAPAAVTSQAQVIKPPL 882  
Qy 153 --SSPSEDDALPSGSPWR-----KKLRKKCEKEKKMEFFPDQDISPLPQ---PSSRN 200  
Db 883 IFVDPNRSFAGPAATPAQAQAASTPRKARASESTARSSSESEDEVDIPATQCLTFGIRT 942  
Qy 201 K-----SRKTEALQKLREVNKRLQDLRLSCLSPKHOSPALQSTDDDEVVLVEGP 249  
Db 943 NVVTMTAHPRIAPKASGASSKSSRISD-----GKKQEGPATQ-----VSKKNP 990  
Qy 250 V-LPQSSRLFTLKIRC-RADLVRLPVRMSEPLQNVVDHMANHLGVSPNRIILLFGESELS 307  
Db 991 ASLPLTQA--ALKVLAQKASEAQPPVARTOPSSGVDVSAVGTLPATSPQ-----STSVQ 1041  
Qy 308 PTATPSTLKLGVADIIDCVVLASSSEATETSQE 340  
Db 1042 AKGTNKLKRPKLPVQQAQKAPESDDSDSDSD 1074

Search completed: January 12, 2003, 10:26:53  
Job time : 20 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 10:25:03 ; Search time 35 Seconds

(without alignments)  
2425,469 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099  
Sequence: 1 MAEPLRGGRPSRGGRGARR.....GKELPADLGLESGLIEVWG 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SEPREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	11 009130	009130 mus musculu
2	1982	99.4	416	11 09CVY5	09CVY5 mus musculu
3	1382	65.8	408	6 09GLZ9	09GLZ9 macaca fasc
4	623	29.7	138	4 096ST9	096ST9 homo sapien
5	154.5	7.4	1217	4 060336	060336 homo sapien
6	145.5	6.9	1072	11 035482	035482 rattus norv
7	140	6.7	1634	5 09XW25	09XW25 caenorhabdi
8	135.5	6.5	2075	13 090WA4	090WA4 tugu rudrip
9	133	6.3	971	5 09XVS4	09XVS4 caenorhabdi
10	131	6.2	664	4 092541	092541 homo sapien
11	130	6.2	704	4 09HSF9	09HSF9 homo sapien
12	129.5	6.2	504	5 018866	018866 caenorhabdi
13	129	6.1	659	5 095XE2	095XE2 caenorhabdi
14	128.5	6.1	484	11 09CS98	09CS98 mus musculu
15	128.5	6.1	635	10 040363	040363 medicago sa
16	128.5	6.1	712	5 061708	061708 caenorhabdi

17	128.5	6.1	755	5 09V4J5	09V4J5 drosophila
18	128.5	6.1	931	13 091995	091995 xenopus lae
19	127.5	6.1	1008	10 094354	094354 oryza sativ
20	127.5	6.1	1167	5 09VN47	09VN47 drosophila
21	127.5	6.1	1372	5 08SX99	08SX99 drosophila
22	127.5	6.1	1408	5 09WOC9	09WOC9 drosophila
23	127.5	6.1	1422	6 09SKU4	09SKU4 canis fam11
24	127	6.1	239	12 057114	057114 muscosa1 dis
25	127	6.1	1368	5 09V6J0	09V6J0 drosophila
26	126	6.0	861	5 021166	021166 caenorhabdi
27	125.5	6.0	1014	11 08VH47	08VH47 mus musculu
28	125.5	6.0	1296	2 08VLP4	08VLP4 streptococ
29	125	6.0	390	5 09XTJ9	09XTJ9 caenorhabdi
30	125	6.0	17352	5 095YM2	095YM2 procambarys
31	124.5	5.9	569	4 096G15	096G15 homo sapien
32	124.5	5.9	670	4 09BM08	09BM08 homo sapien
33	124.5	5.9	757	13 09YHD2	09YHD2 gallus gall
34	124.5	5.9	16	09FB02	09FB02 streptomyce
35	124.5	5.9	1340	16 09LH18	09LH18 streptomyce
36	124.5	5.9	1783	4 015038	015038 homo sapien
37	124.5	5.9	1791	4 060382	060382 homo sapien
38	124.5	5.9	2296	4 09UHA8	09UHA8 homo sapien
39	123.5	5.9	802	4 09UQ37	09UQ37 homo sapien
40	123.5	5.9	815	11 09ERQ2	09ERQ2 rattus norv
41	123.5	5.9	2752	4 09UQ35	09UQ35 homo sapien
42	122.5	5.8	1152	5 09VZ23	09VZ23 drosophila
43	122	5.8	811	5 001905	001905 caenorhabdi
44	122	5.8	876	11 09DC22	09DC22 mus musculu
45	122	5.8	1601	5 09V619	09V619 drosophila

## ALIGNMENTS

RESULT 1	ID	009130	PRELIMINARY;	PRT;	412 AA.
AC	009130;				
DT	01-JUL-1997 (TREMBLrel. 04, Created)				
DT	01-JUL-1997 (TREMBLrel. 04, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	Nuclear protein NIP45.				
GN	NFATC2IP.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97102593; PubMed=8943302;				
RA	Hodge M.R., Chun H.J., Rengarajan J., Alt A., Lieberon R.,				
RA	Glincher L.H.; interleukin-4 transcription potentiated by NIP45.;				
RT	"NF-AT-Driven Interleukin-4 Transcription Potentiated by NIP45.;"				
RL	Science 274:1903-1905(1996).				
DR	EMBL; U76759; AAC52963.1; -				
DR	MGP; MGI:1329015; Nfatc2ip.				
DR	InterPro; IPR000626; Ubiqutin.				
DR	SMART; SM00213; UBO; 1.				
DR	PROSITE; PS50053; UBIQUITIN_2; 1.				
SO	SEQUENCE 412 AA; 45121 MW; DD58FES5C7055C186 CRC64;				
Query Match	100.0%;	Score 2099;	DB 11;	Length 412;	
Best local similarity	100.0%;	Pred. NO.2,3e-140;	Indels	0;	Gaps 0;
Matches 412;	Conservative 0;	Mismatches 0;	Indels	0;	Gaps 0;
QY	1	MAEPLRGGRPSRGGRRGARRGRCPRAROSPARLIPPTVLVDVSDSDEVLVAD	60		
DB	1	MAEPLRGGRPSRGGRRGARRGRCPRAROSPARLIPPTVLVDVSDSDEVLVAD	60		
QY	61	PVEVVALPAPAKPEQSDSDSGAAGPAGPAPPTTLVRRRRRLLDGEGAPVAVYSGK	120		
DB	61	PVEVVALPAPAKPEQSDSDSGAAGPAGPAPPTTLVRRRRRLLDGEGAPVAVYSGK	120		

```

QY 121 VQSSNLIPDNSSLLKLCPEPEDEADLTNSGSSPSDEDDALPSGSPWRKKLRKKCEKEK 180
DB 121 VQSSNLIPDNSSLLKLCPEPEDEADLTNSGSSPSDEDDALPSGSPWRKKLRKKCEKEK 180
QY 181 KMEEFDPQDISPLPQPSSRNKSRKHTEALQKLREVNKRLQDLRSCLSPKQHQSPALQSTD 240
DB 181 KMEEFDPQDISPLPQPSSRNKSRKHTEALQKLREVNKRLQDLRSCLSPKQHQSPALQSTD 240
QY 241 DEVVLVEGVPVLPQSSRLFTLKIRCRADLVRMPSEPLQNVVDHMANHLGVSPNRILL 300
DB 241 DEVVLVEGVPVLPQSSRLFTLKIRCRADLVRMPSEPLQNVVDHMANHLGVSPNRILL 300
QY 301 FGSELSPTATPTSLKLGVAIDIICVVLASSSEATETSQELRLRVQGEKHOMLEISLSP 360
DB 301 FGSELSPTATPTSLKLGVAIDIICVVLASSSEATETSQELRLRVQGEKHOMLEISLSP 360
QY 361 DSPKVLMSHYEAMGLSGHKLSPFFDGTGTLKGKELPADLGESGDLIEVWG 412
DB 361 DSPKVLMSHYEAMGLSGHKLSPFFDGTGTLKGKELPADLGESGDLIEVWG 412

RESULT 2
Q9CVY5 PRELIMINARY; PRT; 416 AA.
AC Q9CVY5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Nuclear factor of activated T-cells, cytoplasmic 2 interacting protein
DE (Fragment).
GN NFAIC2IP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Hoffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya I., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK005947; BAB24331.1; -.
DR MGD; MGI:1329015; Nfatc2ip.
DR InterPro; IPR000626; Ubiquitin.
DR SMART; SM00213; UBQ; 1.
FT NON_TER 1.
SQ SEQUENCE 416 AA; 45551 MW; E4B46F65CC571AF5 CRC64;

Query Match 94.4%; Score 1982; DB 11; Length 416;
Best Local Similarity 97.0%; Pred. No. 4.3e-132;
Matches 394; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

QY 9 GPRSGRGARRA--RGARGCRPRARQSPARLIPDTVLVDVSDSDEEVLEVDAPVEVPV 66
DB 11 GTWSEVPWRPRRSETRGARGCRPRARQSPARLIPDTVLVDVSDSDEEVLEVDAPVEVPV 70

```

```

QY 67 ARLPAPAPQEQSDSDSEGAABGPAGAPRTLVRRRRRLLDPGEAPVVPVYSGKVQSSLN 126
DB 71 ARLPAPAPQEQSDSDSEGAABGPAGAPRTLVRRRRRLLDPGEAPVVPVYSGKVQSSLN 130
QY 127 LIPDNSSLLKLCPSPEDEADLTNSGSSPSDEDDALPSGSPWRKKLRKKCEKEKMEBFP 186
DB 131 LIPDNSSLLKLCPSPEDEADLTNSGSSPSDEDDALPSGSPWRKKLRKKCEKEKMEBFP 190
QY 187 DQDISPLPQPSSRNKSRKHTEALQKLREVNKRLQDLRSCLSPKQHQSPALQSTDDEVVLV 246
DB 191 DQDISPLPQPSSRNKSRKHTEALQKLREVNKRLQDLRSCLSPKQHQSPALQSTDDEVVLV 250
QY 247 EGPVLVLPQSSRLFTLKIRCRADLVRMPSEPLQNVVDHMANHLGVSPNRILLFGESL 306
DB 251 EGPVLVLPQSSRLFTLKIRCRADLVRMPSEPLQNVVDHMANHLGVSPNRILLFGESL 310
QY 307 SPTATPTSLKLGVAIDIICVVLASSSEATETSQELRLRVQGEKHOMLEISLSPSLKV 366
DB 311 SPTATPTSLKLGVAIDIICVVLASSSEATETSQELRLRVQGEKHOMLEISLSPSLKV 370
QY 367 LMSHYEAMGLSGHKLSPFFDGTGTLKGKELPADLGESGDLIEVWG 412
DB 371 LMSHYEAMGLSGHKLSPFFDGTGTLKGKELPADLGESGDLIEVWG 416

RESULT 3
Q9GLZ9 PRELIMINARY; PRT; 408 AA.
AC Q9GLZ9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Hypothetical 44.6 kDa protein.
DE Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]_
RP SEQUENCE FROM N.A.
RT TISSUE=BRAIN PARIETAL LOBE;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050511; BAB17279.1; -.
DR InterPro; IPR000626; Ubiquitin.
DR SMART; SM00213; UBQ; 2.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 408 AA; 44580 MW; 5E0D6DD066AC2F24 CRC64;

Query Match 65.8%; Score 1382; DB 6; Length 408;
Best Local Similarity 70.1%; Pred. No. 9.6e-90;
Matches 293; Conservative 31; Mismatches 78; Indels 16; Gaps 5;

QY 1 MAEPLRGPRSGRGARRARGCRPRARQSPARLIPDTVLVDVSDSDEEVLEV--- 58
DB 1 MAEPLRGPRSGRGARRARGCRPRARQSPARLIPDTVLVDVSDSDEEVLEV--- 52
QY 59 ----ADPVEVPVAPAPAPAPQEQSDSDSEGAABGPAGAPRTLVRRRRRLLDPGEAPV 114
DB 53 ARCAADEVEVAPSEPPGPVASRSDSDSDSEGADARPAGPREPVRRRRLVLDPEAPLV 112
QY 115 PVYSGKVQSSNLIPDNSSLLKLCPEPEDEADLTNSGSSPSDEDDALPSGSPWRKKLRKK 174
DB 113 PVYSGKVQSSNLIPDNSSLLKLYPPGDEEVELADSSGLYHEGSPSP-GSPWKTCLRKK 171
QY 175 CEKEKMEEFDPQDISPLPQPSSRNKSRKHTEALQKLREVNKRLQDLRSCLSPKQHQSP 234
DB 172 -DKEBKKTETIILDLNLSPLSPSPSPRTKGRKTRALKKUSEVKNRKLQDURSCUSPFPQ 230

```

```

QY 235 AAGSTDDVAVVGVVLPSSSLFTIKRCRADVLRLPVPMSEPLQNVVDHMANHLGVSP 294
DB 231 BOOGGEDEVAVVGVVLPSSSLFTIKRCRADVLRLPVPMSEPLQNVVDHMANHLGVSP 290
QY 295 NRILLFGESELPATPTLKLGVADIIDCVVLASSSSATSTSOELRRVQKEXHQML 354
DB 291 SRILLFGESELPATPTLKLGVADIIDCVVLASSSSATSTSOELRRVQKEXHQML 350
QY 355 EISLSPDPLKVLMSHYEAMGLSGHLSFFPDGTYLSKEPLPADLGESGDLIEWWG 412
DB 351 EVSLRDSPLKVLMSHYEAMGLSGHLSFFPDGTYLSKEPLPADLGESGDLIEWWG 408

```

## RESULT 4

```

Q96ST9 PRELIMINARY; PRT; 138 AA.
ID 096ST9
AC 096ST9
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ14639 f16, clone NT2RP2001420, moderately similar to Mus
DE myosin nuclear protein N1P45 mRNA (Hypothetical 15.1 kDa
DE protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Itoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Makamatsu A., Nakamura Y., Nagahari K., Masuo Y., Oshima A.,
RT "NBD human cDNA sequencing project."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Straube R.
RC TISSUE=SKIN;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Straube R.
RC TISSUE=GLIAL TUMOR;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027545; BAB5189.1; -
DR EMBL; BC021551; AAH21551.1; -
DR EMBL; BC018311; AAH18311.1; -
DR InterPro; IPR000626; Ubiquitin.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
KW Hypothetical protein.
SO SEQUENCE 138 AA; 15054 MW; 880930025BC0865D CRC64;

```

```

Query Match 29.7%; Score 623; DB 4; Length 138;
Best Local Similarity 88.4%; Pred. No. 9e-37;
Matches 122; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
QY 275 MSEPLQNVVDHMANHLGVSPNRILLFGESELPATPTLKLGVADIIDCVVLASSSEA 334
DB 1 MSEPLQNVVDHMANHLGVSPNRILLFGESELPATPTLKLGVADIIDCVVLASSSEA 60
QY 335 TETSOELRLRVQKEXHQMLSLSPDPLKVLMSHYEAMGLSGHLSFFPDGTYLSKG 394
DB 61 TETSOELRLRVQKEXHQMLSLSPDPLKVLMSHYEAMGLSGHLSFFPDGTYLSKG 120
QY 395 ELPLADLGESGDLIEWWG 412
DB 121 ELPLADLGESGDLIEWWG 138

```

## RESULT 5

```

O60336 PRELIMINARY; PRT; 1217 AA.
ID O60336
AC O60336;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA0596 protein (Fragment).
GN KIAA0596.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=BRIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998).
CC -1- SIMILARITY: CONTRAINS 7 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; AB011168; BAA25522.1; -
DR InterPro; IPR002114; HPr_Serp_site;
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
DR PROSITE; PS50082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT NON TER 1
SO SEQUENCE 1217 AA; 131098 MW; 0BC4B4C66722BEF5 CRC64;

```

```

Query Match 7.4%; Score 154.5; DB 4; Length 1217;
Best Local Similarity 23.5%; Pred. No. 0.018;
Matches 105; Conservative 46; Mismatches 164; Indels 129; Gaps 21;
QY 6 RGRGRSGRGARARARARARARARARARARARARARARARARARARARARARARAR 65
DB 455 RGRGRSGRGARARARARARARARARARARARARARARARARARARARARARARAR 506
QY 66 -----VARLPAPAKPEODSDSEGAEE-----GPAQARTLVRRRRRL 106
DB 507 ALPLVAKTKKALASVSPALPRSLSHWEMRAQDSVGFLLPAPANPFP---RRGRWV 563
QY 107 DPGEAPVVPVYSGVQSSILNIPNSSLKLCPS--EP-EDPADLTNCGSSPSEDDALPS 163
DB 564 QPG-----VELSVRSMIDLRLQLETLAPSLQDPSQDSLAIIPGPRKHQGEALET 612
QY 164 G-----SPARKKLRKCC-----EKKEP-----KKMEEPDODISLPDPS--SNK 201
DB 613 SLTSGNEKPRPQASQPCSPYHILRLSQEGVFAODLEPAIEIGIYPPPSDNPPTMDT 672
QY 202 SRKTHEALQK-----LREVNKRLOD-----LRSCSPKQHQSPALQSTDDDEVVL 245
DB 673 SEFOVAPARQTLGRVYPGSSSEKSPDACSVDYSSCLSSPH--PTDSESTEPFLS 730
QY 246 VEGVLPQSSRLFTLKIRCRADVLRLPVPMSEPLQNVVDHMANHLGVSPNRILLFGESE 305
DB 731 VDQ-----ISSDLDEPAEGDEEBEBSGKMP-----YGIQE 762
QY 306 LSPATP-----STLKLGVADIIDCVVLASSSEATSTSOELRLRVQKEXHQML 355
DB 763 GSP-QTPDQEOFLKQHFETTLASGAAPAPV-QVPERSESRSISSRFLQVOTRPLREP-- 818
QY 356 ISLSPDPLKVLMSHYEAMGLSGH 381
DB 819 ---SPSSSLALMSRPAQVPAQSGEQ 841

```



OK NCBI\_TaxID=31033;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVARY;  
 RA Bolland D.J., van Geel M., Carim Todd L., Beck A.F., Grewal P.K.,  
 RA van der Maarel S., Frants R.R., de Jong P.J., Hewitt J.E.;  
 RT "Comparative sequence analysis of an evolutionary chromosomal  
 RT breakpoint in human, mouse and Pugu rubripes."  
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF389403; AAK69357.1; -  
 SO SEQUENCE 2075 AA; 229612 MW; 88A27A058F15045B CRC64;  
  
 Query Match 6.5%; Score 135.5; DB 13; Length 2075;  
 Best Local Similarity 22.4%; Pred. No. 0.78;  
 Matches 92; Conservative 48; Mismatches 125; Indels 145; Gaps 19;  
  
 QY 13 RCGRARRRARGRCRCRARGSPARL-----IPDTV-----LVLDVSDSDEVLVAD 60  
 DB 652 RCGNSNNQGRINTECEINNRSAANLRSLNTSTTECOYNRDTRYNEVDEDEDVLDNEE 711  
 QY 61 PVEVVARLPAPAKPEQSDSDSEGAAEGPAGAPRTLVRRRRLDLPGEAPVPVYSGK 120  
 DB 712 GAR-----AAAAP-----DSEGS-----RSSLGNAGEF-VHKYHQCK 744  
 QY 121 VOSSLNLIPDNSSLKLCPSPEDEADLTNCGSSPEEDAL-----PSGSPWRKK 170  
 DB 745 VKOKRLOLQELMAMVQ-----SDTD-----GTANEDDILHQPNNTRAGPKSPREP 794  
 QY 171 LAKKCEKEKMEEFPPDODISPLPQSSRNKRKHTKALQKREVNKRLODRLSCISPKQ 230  
 DB 795 LSSKA--RKLYEE-----KIROQKOLKHLDEKRLTELOGKIRLDQ 836  
 QY 231 HOSPALQST-----DDEVLVVEGPVLPQSSRLFTLKIRGRADLVLRVPMSEPLQNVVDH 285  
 DB 837 WACPDQSSVCSTSGSGGLRKVPV-----TPVSIPPGPVQAVAS 876  
 QY 286 MANHLGVSPNRIILLFGESELSPTATPTLKGVADIDICVLIASSSEATSTSOELR--- 342  
 DB 877 -----GPKKNVSL-----KKAAPPAVAVPAD-----NELMSEMRHQ 910  
 QY 343 -LRVQKEKHQMLETSL-----PDSPLKVLMSHYEAMGLSGHKLS 383  
 DB 911 TLREBLRHRKHLLEYLIAHQRRSGPTDSP-----RHNRGSLASPSHTVS 955  
  
 RESULT 9  
 Q9XVS4 PRELIMINARY; PRT; 971 AA.  
 AC Q9XVS4;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE C25A1.10 protein.  
 GN C25A1.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodetinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Mortimore B.J.;  
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.  
 RN (2)  
 RP MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018 (1998).  
 DR EMBL: Z81038; CAB02755.1; -  
 DR InterPro: IPR002965; P\_tich\_exten.  
 DR PRINTS: PRO1217; PRICHEXTENSN.  
 SO SEQUENCE 971 AA; 101097 MW; 832936115B58709B CRC64;

Query Match 6.3%; Score 133; DB 5; Length 971;  
 Best Local Similarity 23.3%; Pred. No. 0.44;  
 Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;  
  
 QY 29 PPARQSPARLLPDTVLVDLV-----DSDEVLVADPVEVVARLPAPAKPEDSDSDSG 85  
 DB 231 PAAKTPPAKTPKPPVAKKAESSSDSDDEKKPVAKP--APAKATPKPAKKADSSSDSD 288  
 QY 86 ----AAEGPAGAPRTLVRRRRLD-----PGAPVPVYSGVQSSLNLI 128  
 DB 289 DEAPAKTPPAKAPKPVAKKAESSSDSDDEKKPAKTPPAKATPKPAKKAESSSDSD 348  
 QY 129 PDNSSLKLCPS--PEDEADLTNCGSSPEEDALPSGSPW-RKKLRKKCEKMEEF 185  
 DB 349 DEKKVPAPPAKATPKPVAKKAESSSDSDDEKKPAKTPPAKATPKPVAKKAESSSDS 408  
 QY 186 PDQDISPLPQSSRNKS-----RKHTALOKLREVNKRLODLNSC 225  
 DB 409 SDDEKKPVAKPTSAKATPKPAKKAADSSSDSDDEAPAKTPPAKAPPAKKA--SS 465  
 QY 226 LSPKHOSPALQSTDEV 243  
 DB 466 SDSDEKPAKSTPAKI 483  
  
 RESULT 10  
 Q92541 PRELIMINARY; PRT; 664 AA.  
 AC Q92541;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE KIAA0252 protein (Fragment).  
 GN KIAA0252.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BONE MARROW;  
 RA MEDLINE=97191544; PubMed=9039502;  
 RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Chara O.,  
 RA Tanaka A., Kotani H., Miyajima N., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. VI.  
 RT the coding sequences of 80 new genes (K1A0201-K1A0280) deduced by  
 RT analysis of cDNA clones from cell line KG-1 and Brain."  
 RL DNA Res. 3:321-329 (1996).  
 DR EMBL: D87440; BAA13382.1; -  
 DR InterPro: IPR004343; Plus-3.  
 DR Pfam: PF03126; Plus-3; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SO SEQUENCE 664 AA; 75809 MW; 2C6537A98BFADEB CRC64;  
  
 Query Match 6.2%; Score 131; DB 4; Length 664;  
 Best Local Similarity 21.5%; Pred. No. 0.37; Indels 148; Gaps 19;  
 Matches 101; Conservative 69; Mismatches 151;  
  
 QY 37 RLIPDTVLVDLVSDS--DEVLVLA-----DPVEVVARLPAPAKPEQSDSDSEGAA 87  
 DB 2 RVIDSDTDEDSGSDNLQDELLSLAKRKRSDSEKEPVSQ-PAASDSESTSDSDDEWT- 59  
 QY 88 ECPGAPRTVLRRRRRLDLPGEAPVAVVYSGKVOSSLNLIPDNSSLKLCPSPEDEAD 147  
 DB 60 ---FGSNKKKKKAKRKLKKG--TMKKQANKTASSGSDKD--SSASSAPAEGEVSDS 112  
 QY 148 LTNCGSSPEEDALPSGSPW----- 167  
 DB 113 DSNSSSSSDSDSDEDFHDGVEDLMGDEBRARLEQMTKEREDELFRIRKREVL 172  
 QY 168 -----RKLR--KKCEKMEEFPPQDISPL-----PPSSRNKR-----KH 205

```
Db 173 KRPFBIKKLTKAKKKEKKKQEQEKKLTOIQESQVTSNKKERRSKDEKLDKK 232
Qy 206 TEALQKLR-EVNRKLODLRSLCPKQ-----HOSPALQSTD 240
Db 233 SQAMEELKAEREKKRNTAELLAKQPLKTSEVYSDDEEEEDDKSSEKSDRSRTSSD 292
Qy 241 DEVVLVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSPNRIILL 300
Db 293 EEEKEEIP--PKSQ-----PVSLEPEL-----NVRLSRHKLRLRW 326
Qy 301 FGESELSPTATPSTLKL-----VADIIDCVLASSSE--ATETSOELRLRVQ 347
Db 327 CHMPFFAKTVTGCFVIRIGNHNHSPVYVAEITGVETAKVYQLGCTRTNKLQLRHGN 386
Qy 348 KEKHQMLEISLSDPSLKVLMVSHYEBAMGLSGHKLGFDFGTGKLSQKEL 396
Db 387 DQVRFRLEFVSNGEFTSEFPMK-WKEAMFSAGNQLPTL---DEINKKEL 431

RESULT 11
Q9H5F9 PRELIMINARY; PRT; 704 AA.
AC Q9H5F9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-DEC-2001 (TREMBlrel. 16, Last sequence update)
DE CDNA: FLJ23471 fis, clone HS111969.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL "NEO human cDNA sequencing project.";
RL EMBL; AK027124; BAB15667.1; -.
DR InterPro: IPR002965; P rich extensin.
DR PRINTS; PR01217; PRICHEXTENSIN.
SQ SEQUENCE 704 AA; 75768 MW; 858F94EEA2C1F8C6 CRC64;

Query Match 6.2%; Score 130; DB 4; Length 704;
Best Local Similarity 22.6%; Pred. No. 0.47;
Matches 100; Conservative 63; Mismatches 173; Indels 106; Gaps 20;

Qy 10 PRSRGGRARRARGRCPRARQSPARLIPDTVLVDVSDSEEVLEVADPV----- 62
Db 286 PKTEAPQASPLAKPLOSSSPVLGLPSRMEPPAPL---STSTSOASALPPAGRRLAE 341
Qy 63 -----EVPVRLPAPAKPEQSDSDSEGAEGPAG-----AP 94
Db 342 SSGVGRVAGSRKPAPMAKGSITLTQDMSTSLQEGDGPAGWRANLKPVDRRSPAE 401
Qy 95 RTLVRRRRRLDP--GEAP--VVPVYSGKVQSSLLNI-PDNS-----SLLKLCPS 141
Db 402 RTLKPEPRALAEPRAGEAPRKVSGFAGSVHITLTPVRDTPRPASPGPSLPASPS 461
Qy 142 PEDEADLTNSGSPSDEDDALPGSPWRKRLKKCEKEEK--MEEPDQDISP-----LP 194
Db 462 PR--RRRLAVPASLDVCDNWLRLPPEPPQEARVQSWKEEKPHLQGRPLSPANVPALP 520
Qy 195 QPSSRNKSRKHTALOKLREVNRKLODLRSLSPKQHSQALQST-----DD--EVLVL 246
Db 521 GETVTSPVRLHPDLYSP-EIQQLQDIERRLDALERGLVEKRLRAAEGDAEDSLMW 579
Qy 247 EGPVLPSSRLFTLKIRCRADLV--RLPVRMSEPLQNVVDHMANHLGVSPNRIILL---FG 302
Db 580 DWFNLTHEKQLL---LRQESLWYKSAQRLBEQLDIEGELRLMAKPEALKSLQERRR 636
Qy 303 ESELSPTATPSTLKLGVADIIDCVLASSSEATETSOELRLRVQKKEKHQMLEISLSPDS 362
```

```
Db 637 EQEL-----LEQVSTVND-----RSDIVSLDDRLREQ--BEDQMLRMI----- 676
Qy 363 PLKVLMSHYEAMGLSGHKLSP 384
Db 677 -----EKJGLQKSKSKF 688

RESULT 12
Q18866 PRELIMINARY; PRT; 504 AA.
AC Q18866;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical 55.4 kDa protein.
GN C55C3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodirinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2; PubMed=9851916;
RX MEDLINE=99069613;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Woessne J., Stellyes L.;
RT "The sequence of C. elegans cosmid C55C3.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53335; AL27231.1; -.
KW Hypothetical protein.
SQ SEQUENCE 504 AA; 55405 MW; 9605686914F44659 CRC64;

Query Match 6.2%; Score 129.5; DB 5; Length 504;
Best Local Similarity 21.3%; Pred. No. 0.33;
Matches 91; Conservative 52; Mismatches 127; Indels 157; Gaps 21;

Qy 11 RSRGGRARRARGRCPRARQSPARLIPDTVLVDVSDSEEVLEVADPVVPARLP 70
Db 29 RRRGQNQIKKKPAK---PLKRITP-----NNSDEKLKKMKSP----- 63
Qy 71 APAKEQSDSDSEGAEGPAGAPRTLVRRRRRRLDP---GEAPVVPYSGKVQSSLLNL 127
Db 64 -PKPLQATDSD-----ILQPLSNAMTVP-----ESKESANR 96
Qy 128 IPDNSSL-----LKLCPSEPE-----DEADLTNSGSPSDEDDALPGSPWRKCLR 172
Db 97 SQSSKINLNTDLDKDC-SEARAIPKCDKLTDFQSTTNSQSPA--DAISAVAP-TKLTG 152
Qy 173 KKCEKEEKMEFPD-----QDI-----SPLQPSSRNK-----SRK 204
Db 153 SOSQSERQIEKMPNSFRELODSKFKPAEAPIPKPEERKSGSTEQSTTPTKLAGSQSPS 212
Qy 205 HTEALQKLRVNRKLOD--LRSLSPKQHSQALQSTDD----- 241
Db 213 ERQEIQQMPNSFRELODSKFKPAPSPKSEKGAQSLSEPPSVPGNNRRNSTSQVDITIS 272
Qy 242 ----EVLVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDH-----MANHLG 291
Db 273 PVPYTKLVGTQSPSERQIEQMPNSFRELODSKFKPAQAPNPKPERVERGSAEQSMSEPLS 332
```

QY	292	VSPRRILLFESELSPTATPTSLK-----LGVAADIICVLA-----SSSEA	334
		::          ::          ::          ::          ::	
Db	333	IS---KRAFG-SPIAPKRRSPLOAPULETLAPPTIDAPTAIETATRSAPFSSHS	387
QY	335	TETSOEL	341
		:	
Db	388	MDPNSNL	394

RESULT 13			
095XE2			
ID	Q95XE2	PRELIMINARY;	PRT; 659 AA.
AC	Q95XE2;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Hypohectical 71.7 kDa protein.		
GN	Y73B3A.1.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea		
OC	Rhabditidae; Pelodertinae; Caenorhabditis.		
OX	NCBI_TaxId=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RC	STRAIN=BRISTOL N2;		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	None;		
RT	"genome sequence of the nematode C. elegans: a platform for		
RT	investigating biology. The C. elegans Sequencing Consortium.";		
RL	Science 282:2012-2018(1998).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RC	STRAIN=BRISTOL N2;		
RA	Leonard S.;		
RT	"The sequence of C. elegans cosmid Y73B3A.";		
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RC	STRAIN=BRISTOL N2;		
RA	Waterston R.;		
RT	"Direct Submission.";		
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.		
RL	EMBL:AC084159; AAK39359.1; --		
KM	Hypohectical protein.		
QO	SEQUENCE 659 AA; 347A29640289E1B8 CRC64;		

```

Qy      353 MEISLSPDSPPLKVLMSHYEEAMGLSGHKISFF 3855
          : | : | : | : | : |
Db      611 V-----DDAPA-----EHSMSLQGRVLTMF 6300

```

RESULT	14
QC9CS98	
ID	QC9CS98
AC	QC9CS98;
DT	01-JUN-2001 (TREMBlrel_17, Created)
DT	01-JUN-2001 (TREMBlrel_17, Last sequence update)
DT	01-JUN-2001 (TREMBlrel_17, Last annotation update)
DE	3930402D05Rik protein (Fragment).
CN	3930402D05Rik.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	[NCBI_Taxid=10090];
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shinnagawa A., Shibata K., Yoshino M., Ishii Y.,
RA	Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuoka S.,
RA	Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Peole G., Quackenbush J.,
RA	Schirai L.M., Staubli F., Suzuki R., Tomita M., Wagner T., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Gustincich M.-J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
RA	Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,
RA	Hayashizaki Y.;
RT	"Functional annotation of a full-length mouse cDNA collection";
RL	Nature 409:685-690(2001).
DR	EMBL: AK014457; BA829364.1; -
DR	MED; MGI:1921269; 3930402D05Rik.
QO	NON_TER
SEQUENCE	484 AA; 55413 MW; 328371FF5211F54 CRC64;





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 10:23:09 ; Search time 39 Seconds

(without alignments)  
1407.673 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099

Sequence: 1 MAEPLRGGRGRRGARR.....GKELPADLGESGDLIEWG 412

Scoring table: BLOSUM62

Searched: Gap0 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : A\_Geneseq/101002.\*

Listing first 45 summaries

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2099	100.0	412	18	AAW34091	Mouse NF-AT intera
2	2099	100.0	412	20	AAV08331	Mouse NIP45 protei
3	1402.5	66.8	419	23	AAU10376	NF-AT interacting
4	1389.5	66.7	419	20	AAV08330	Human NIP45 protei
5	1399.5	66.7	419	22	AAW93780	Human polypeptide,
6	1079	51.4	303	23	AAU10374	NF-AT interacting
7	987	47.0	334	23	AAU10372	NF-AT interacting
8	785.5	37.4	286	23	AAU10373	NF-AT interacting
9	623	29.7	138	22	AAW94004	Human stomach can
10	623	29.7	138	22	AAW94289	Human protein sequ

11	255	12.1	55	23	AAU10375	NF-AT interacting
12	235	11.2	87	23	AAO14942	Human NF-AT intera
13	154.5	7.4	1219	22	ABG12658	Novel human diago
14	154.5	7.4	1279	22	AAW78969	Human protein SEO
15	154.5	7.4	1301	22	AAW79953	Human protein SEO
16	148	7.1	710	22	ABG20363	Novel human diago
17	143.5	6.8	2013	22	ABG12657	Novel human diago
18	137	6.5	990	23	AAE22762	Rat PGC-3 protein.
19	134	6.4	346	23	ABW06133	Human NS protein s
20	132	6.3	586	18	AAW10423	Cytomegalovirus UL
21	128.5	6.1	755	22	ABW59259	Drosophila melanog
22	128	6.1	281	22	ABG05583	Novel human diago
23	128	6.1	281	22	ABG18846	Novel human diago
24	127.5	6.1	1167	22	ABW57757	Drosophila melanog
25	127.5	6.1	1408	22	ABW58704	Drosophila melanog
26	127	6.1	1368	22	ABW60262	Drosophila melanog
27	126.5	6.0	676	22	AAW47218	Human NOV9 protein
28	126.5	6.0	676	22	AAW09441	Human SB-HDGF prot
29	126.5	6.0	676	22	AAW64373	Human hepatome cel
30	126.5	6.0	676	23	ABW05039	Human NOV4a protei
31	126.5	6.0	717	22	AAE09442	Human SB-HDGF prot
32	125.5	6.0	1014	23	AAU96745	PGC2-PPAR gamma co
33	124.5	5.9	518	21	AAW53332	Human colon cancer
34	124.5	5.9	555	22	AAW35073	Human colon cancer
35	124.5	5.9	671	21	AAW9426	Human Prol604 (UNQ
36	124.5	5.9	671	22	AAU29196	Human PRO polypept
37	124.5	5.9	671	22	AAW61175	Protein of the inv
38	124.5	5.9	671	23	ABW05042	Human NOV4d protei
39	124.5	5.9	1783	23	ABW97448	Novel human protei
40	124	5.9	1191	22	AAW80219	Human protein SEO
41	123.5	5.9	671	23	AAW97038	Human LP protein L
42	123	5.9	950	20	AAW33298	Human membrane spa
43	122.5	5.8	1152	22	ABW57951	Drosophila melanog
44	122.5	5.8	1152	22	ABW66110	Drosophila melanog
45	122.5	5.8	1152	22	ABW67224	Drosophila melanog

ALIGNMENTS

AAW34091	RESULT 1
AAW34091	AAW34091 standard; Protein; 412 AA.
AAW34091	AAW34091;
DT	18-MAY-1998 (first entry)
DE	Mouse NF-AT interacting protein 45.
XX	NF-AT Interacting Protein 45; NIP45; yeast two-hybrid assay; mouse;
KW	Rel homology domain; RHD; T cell; transcription factor; cancer;
KW	interleukin-4; IL-4; development; Th1; Th2; cytokine; allergy;
KW	autoimmune disease; transplantation.
XX	Mus sp.
OS	
XX	
FT	Key
FT	Region
FT	Location/Qualifiers
FT	6..37
FT	/note= "highly basic region"
PN	W09739721-A2.
PD	30-OCT-1997.
PF	23-APR-1997;
PF	97WO-US06708.
PR	25-NOV-1996;
PR	96US-0755592.
PR	23-APR-1996;
PR	96US-0636602.
PR	25-NOV-1996;
PR	96US-0755584.
PA	(HARD ) HARVARD COLLEGE.
XX	

PI Glincher LH, Ho I, Hodge MR;  
XX WPI; 1997-535556/49.  
DR N-PSDB; AAT93045.  
XX Production of cytokine(s) associated with Th2-type helper T cells -  
PT particularly for controlling development of Th1 and Th2 cells for  
PT treatment of allergy, autoimmune disease etc.  
XX  
PS Claim 31; Fig 11; 151pp; English.  
XX  
CC This is the amino acid sequence of the mouse NF-AT Interacting Protein 45  
CC (NIP45). The gene sequence was isolated by using a yeast two-hybrid  
CC detection system for proteins that interact with the NF-AT Rel homology  
CC domain (RHD). The assay used, as a "bait", a 900 bp fragment of the  
CC murine NF-ATp encoding the region spanning amino acids 228-250. NF-AT is  
CC a multisubunit transcription complex containing a cyclosporin A sensitive  
CC cytoplasmic phosphoprotein and an inducible component of the AP-1 family  
CC of transcription factors. The screen was carried out on a cDNA library  
CC prepared from the murine T cell line D10. One class of proteins,  
CC designated NIP45, bound the NF-AT-RHD region with high affinity. NIP45  
CC can be used in a claimed method to inhibit or stimulate production of  
CC NF-AT family protein, particularly interleukin-4 (IL-4) and the treated  
CC cells may be administered to control development of Th1 or Th2 cells by  
CC modulating the production of a T-helper type 2 associated cytokine.  
CC Especially the method is useful in the inhibition of Th2 in allergy,  
CC cancer or infections, and promotion of Th2 in autoimmune disease and  
CC transplantation.  
XX  
SQ Sequence 412 AA;  
  
Query Match 100.0%; Score 2099; DB 18; Length 412;  
Best Local Similarity 100.0%; Pred. No. 1.5e-174;  
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDEEVLEAD 60  
Db 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDEEVLEAD 60  
  
QY 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120  
Db 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120  
  
QY 121 VOSSLNLIIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180  
Db 121 VOSSLNLIIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180  
  
QY 181 KMEFPDQDISPLPQSSRNKSRKTEALQKLRVKNRLQDLRSLSPKQHSQSPALQSTD 240  
Db 181 KMEFPDQDISPLPQSSRNKSRKTEALQKLRVKNRLQDLRSLSPKQHSQSPALQSTD 240  
  
QY 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300  
Db 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300  
  
QY 301 FGSELSPTATPSTTLKLGVDIICVVLASSSEATSTSOELRLRVQKQKQHMLEISLSP 360  
Db 301 FGSELSPTATPSTTLKLGVDIICVVLASSSEATSTSOELRLRVQKQKQHMLEISLSP 360  
  
QY 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412  
Db 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412  
  
RESULT 2  
AAY08331  
ID AAY08331 standard; Protein; 412 AA.  
XX  
AC AAY08331;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Mouse NIP45 protein.

XX NIP45; mouse; transcription transactivator; IL-4; interleukin-4;  
KW NF-AT interacting protein-45; modulator; treatment; inflammation;  
KW autoimmune disease; HTP screening; drug testing; allergic; therapy;  
KW T-cell dependent response; cytokine; diagnostic; immunosuppressant.  
XX  
OS Mus musculus.  
XX  
PN WO9921993-A1.  
XX  
PD 06-MAY-1999.  
XX  
PF 21-OCT-1998; 98WO-GB03141.  
XX  
PR 24-OCT-1997; 97GB-0022388.  
XX  
PA (ZENE ) ZENECA LTD.  
XX  
PI Liu D, Zhao J, Zhou H;  
XX  
DR WPI; 1999-312964/26.  
XX  
PT Polynucleotides encoding human NF-AT interacting protein (NIP45)  
XX  
PS Disclosure; Fig 4; 84pp; English.  
XX  
CC This invention describes a novel human NF-AT interacting protein-45  
CC NIP-45. Human NIP45 is a transcriptional trans-activation factor of the  
CC interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used  
CC to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or  
CC autoimmune disease. NIP45 is a good candidate target for HTP screening  
CC and/or testing system for drugs that will alleviate T-cell dependent  
CC autoimmune and allergic responses, and for cytokine-based therapies of  
CC chronic disease. Expression of NIP45 can be inhibited, and IL-4  
CC expression modulated, by administering an antisense molecule. Antibodies  
CC against NIP45 and primers can be used in diagnostic assays. Discovery of  
CC a human NIP45 is advantageous in that it provides the ability to control  
CC IL-4 transcription, which is of importance for anti-inflammation and  
CC immunosuppressant drug development.  
XX  
SQ Sequence 412 AA;  
  
Query Match 100.0%; Score 2099; DB 20; Length 412;  
Best Local Similarity 100.0%; Pred. No. 1.5e-174;  
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDEEVLEAD 60  
Db 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDEEVLEAD 60  
  
QY 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120  
Db 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120  
  
QY 121 VOSSLNLIIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180  
Db 121 VOSSLNLIIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180  
  
QY 181 KMEFPDQDISPLPQSSRNKSRKTEALQKLRVKNRLQDLRSLSPKQHSQSPALQSTD 240  
Db 181 KMEFPDQDISPLPQSSRNKSRKTEALQKLRVKNRLQDLRSLSPKQHSQSPALQSTD 240  
  
QY 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300  
Db 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300  
  
QY 301 FGSELSPTATPSTTLKLGVDIICVVLASSSEATSTSOELRLRVQKQKQHMLEISLSP 360  
Db 301 FGSELSPTATPSTTLKLGVDIICVVLASSSEATSTSOELRLRVQKQKQHMLEISLSP 360  
  
QY 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412  
Db 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412

RESULT 3  
 AAU10376 standard; Protein: 419 AA.  
 AC AAU10376;  
 DT 14-FEB-2002 (first entry)  
 DE NF-AT interacting protein 45, NIP45.  
 XX NF-AT interacting protein 45 variant; NIP45V; immunosuppressive; human;  
 KM antiapoptotic; antidiabetic; antitumor; dermatological; antiinflammatory;  
 KM antiproliferative; antibacterial; vasodilator; autoimmune; allergy; asthma;  
 KM nuclear factor of activated T cells; chronic inflammatory disease;  
 KM allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;  
 KM vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;  
 KM diarrhetic colitis; polymyositis; lymphofollicular thyroiditis; gout;  
 KM dermatomyositis; Sjogren's syndrome.  
 XX Homo sapiens.  
 OS  
 XX MO200181574-A2.  
 XX 01-NOV-2001.  
 PD 25-APR-2001; 2001MO-EP04635.  
 PF 25-APR-2001; 2000US-199356P.  
 PR 25-APR-2001; 2000US-199356P.  
 XX (FARB ) BAYER AG.  
 PA  
 PI Encinae J;  
 XX  
 PI WPI; 2002-055355/07.  
 DR N-PSDB; AAS17539; AAS17540.  
 XX  
 PT Novel nuclear factor of activated T cells interacting protein NIP45  
 PT variants which can be regulated for treating conditions involving  
 PT inflammatory processes such as allergy, asthma and autoimmune diseases  
 PT  
 XX  
 PS Disclosure; Fig 12; 107pp; English.  
 XX  
 CC The invention relates to an isolated nuclear factor of activated T cells  
 CC (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and  
 CC polynucleotide (II) encoding (I). (I) and (II) are useful for screening  
 CC for agents which decrease the activity of NIP45V polypeptide. (I) is also  
 CC useful for screening for agents which regulate the activity, preferably  
 CC cyclic AMP formation, mobilisation of intracellular calcium or  
 CC phosphoinositide metabolism of human NIP45V. (II) is useful for detecting  
 CC a polynucleotide encoding a NIP45V polypeptide in a biological sample, by  
 CC hybridising the polynucleotide to a nucleic acid material of a biological  
 CC sample. Antibody which specifically binds to (I) or (II) is useful for  
 CC reducing the activity of NIP45V in a cell, and for detecting (I) in  
 CC a sample. (I) is useful for modulating the activity of NIP45V in a  
 CC disease, such as autoimmune, allergic, infectious or chronic inflammatory  
 CC disease or asthma. NIP45V is useful for controlling diseases including  
 CC allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal  
 CC catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,  
 CC diarrhetic colitis, systemic inflammatory response syndrome, polymyositis,  
 CC lymphofollicular thyroiditis, sepsis, dermatomyositis, mixed connective  
 CC tissue disease (MCTD), Sjogren's syndrome and gout. (I) is also useful  
 CC in diagnostic assays for detecting diseases and abnormalities or  
 CC susceptibility to diseases related to the presence of mutations in  
 CC nucleic acid sequence which encode (I). (I) or its fragment is useful  
 CC in raising specific antibodies. The present sequence represents the  
 CC amino acid sequence of NIP45 as described in the invention.  
 XX  
 SQ Sequence 419 AA;

Query Match 66.8%; Score 1402.5; DB 23; Length 419;

Best Local Similarity 70.1%; Pred. No. 9.5e-114;  
 Matches 295; Conservative 30; Mismatches 85; Indels 11; Gaps 4;  
 QY 1 MAEPLRGGRSPRG---CRGARBARGARGRCPRAROSPRLIPDTVLVDVSDSEEVLE 57  
 Db 1 MAEPVGRKGRMSGSGARGRGGRGGRAPRAORSSRGTLDVVSDVLDVSDSEEVLE 60  
 QY 58 V-----ADPVEVPVAPRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLLDPGEA 111  
 Db 61 VATARGADEVEVEPEPPPGPVASRDNSDSEGEDRRPAGPPRPVRRRRLLVDPGEA 120  
 QY 112 PVVPYSGKVGSSNLTLIDNSSLKLCSEBPEDEADLINSGSSPEEDALPSGSPWRKL 171  
 Db 121 PLVPYSGKVGSSNLTLIDNSSLKLCSEBPEDEADLINSGSSPEEDALPSGSPWRKL 179  
 QY 172 RKCKEKKKKKEEFPDODISPLPPSSRNKRKTEALOKLREVKRRLODLRSCSPKQH 231  
 Db 180 RTK-DKEKKKTEFLDLDNSPLSPSPPTKSRTRTRALKLSEVVKRRLODLRSCSPKRP 238  
 QY 232 QSPALQSTDEVEVLVEGVLPPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHANHLG 291  
 Db 239 QGQEQGQGEDEVVLVEGVLPPETPLPRLPLKIRCRADLVRLPLRMSEPLQSVVDHANHLG 298  
 QY 292 VSPNRILLPGSESLPATPSTLKLGVADIIDCVVLASSSEATETSGELRLRVGCKEK 351  
 Db 299 VSPSRILLPGSESLPATPSTLKLGVADIIDCVVLASSSEATETSGELRLRVGCKEK 358  
 QY 352 QMLRSLSPDSPSLKVLMSHYEAMGLSGHKLSPFPDGTKLSGKELPAULGLSSGDLIEW 411  
 Db 359 QTLVSLSRDSEPLKTLMSHYEAMGLSGHKLSPFPDGTKLSGELPAULGLSSGDLIEW 418  
 QY 412 G 412  
 Db 419 G 419

RESULT 4  
 AA08330  
 ID AA08330 standard; Protein: 419 AA.  
 AC AA08330;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE Human NIP45 protein.  
 XX  
 KM NIP45; human; transcription transactivator; IL-4; interleukin-4;  
 KM NF-AT interacting protein-45; modulator; treatment; inflammation;  
 KM autoimmune disease; HTP screening; drug testing; allergic; therapy;  
 KM T-cell dependent response; cytokine; diagnostic; immunosuppressant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9921993-A1.  
 PD 06-MAY-1999.  
 XX  
 PF 21-OCT-1998; 98MO-GB03141.  
 XX  
 PR 24-OCT-1997; 97GB-0022388.  
 DR N-PSDB; AAX56955.  
 XX  
 PT Polynucleotides encoding human NF-AT interacting protein (NIP45)  
 XX  
 PS Claim 1; Fig 3; 84pp; English.  
 XX  
 CC This invention describes a novel human NF-AT interacting protein-45

CC NIP-45. Human NIP45 is a transcriptional trans-activation factor of the  
CC interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used  
CC to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or  
CC autoimmune disease. NIP45 is a good candidate target for HTP screening  
CC and/or testing system for drugs that will alleviate T-cell dependent  
CC autoimmune and allergic responses, and for cytokine-based therapies of  
CC chronic disease. Expression of NIP45 can be inhibited, and IL-4  
CC expression modulated, by administering an antisense molecule. Antibodies  
CC against NIP45 and primers can be used in diagnostic assays. Discovery of  
CC a human NIP45 is advantageous in that it provides the ability to control  
CC IL-4 transcription, which is of importance for anti-inflammation and  
CC immunosuppressant drug development.

XX  
SQ Sequence 419 AA;

Query Match 66.7%; Score 1399.5; DB 20; Length 419;

Best Local Similarity 70.1%; Pred. No. 1.7e-113; Mismatches 85; Indels 11; Gaps 4;

Matches 295; Conservative 30; Mismatches 85; Indels 11; Gaps 4;

QY 1 MAEPLRGGRPRSRG---GRGARRARGRCPRARQSPARLIPDTVLVLDVSDSDEEVLE 57

DB 1 MAEPVKGKRWGSGGAGRGGRGGRGRRPRAQSPSRGTLDVVSDVLTDSDEILE 60

QY 58 V-----ADPVEVPVARLPAPAKPEQSDSDSEGAAGGAGAPRTLVRRRRRLDPGEA 111

DB 61 VATARGADEVVEPEPPGPGPVASRDNSDSEGEDRRPAGPPREPVRRLVLDPGEA 120

QY 112 PVVPVYSGKVOSSLNLIIPDNSSLLKCPSEPEDEADLTNSGSSPSEDALPSGSPWRKKL 171

DB 121 PLVPVYSGVKVKSRLIPDDLKLLKYPGDEEAEALADSSGLYHEGSPSP-GSPWKTKL 179

QY 172 RKCKEKEKMEFFPDQDISPLPQSSRNKSKKHTKALOKLREVNRKLDLRLVQCKEKH 231

DB 180 RTK-DKEEKKTEFLDNLSPSPSPRPTKSRTHTRALKKLVSEVNRKLDLRLVQCKEKH 238

QY 232 QSPALQSTDEVVLVEGVPVLPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 291

DB 239 QGQEQQOQDEVVVLVEGPTLPETPRLFKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 298

QY 292 VSPNRILLFGESELSPATPTPSTLKLGVADIIDCVVLASSSEATETSOELRLVQCKEKH 351

DB 299 VSPSRILLFGETELSPATPTPRTLKLGVADIIDCVVLTSSEATETSSQQLQVRQCKEKH 358

QY 352 QMLEISLSPDSPKVLMSHYEAMGLSGHKLSFFFDGTTKLSGKELPADLGSGDLEIYVW 411

DB 359 QTLVSLSRDSPKLTLMSHYEAMGLSGKLSFFFDGTTKLSGRELPA DLGMEGSDLEIYVW 418

QY 412 G 412

DB 419 G 419

RESULT 5

AAU93780

ID AAU93780 standard; Protein; 419 AA.

XX

AC AAU93780;

XX 06-NOV-2001 (first entry)

DT Human polypeptide, SEQ ID NO: 3794.

DE Human; full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

OS EP1130094-A2.

XX 05-SEP-2001.

PD 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX

PR

PR 11-JAN-2000; 2000JP-0118774.  
PR 02-MAY-2000; 2000JP-0183765.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX WPI; 2001-524255/58.  
DR N-PSDB; AAK94735.

XX 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation -  
PT  
XX

PS Claim 8; SEQ ID NO 3794; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesising full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
CC molecules have been determined. Primers for synthesising the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a polypeptide  
CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 419 AA;

Query Match 66.7%; Score 1399.5; DB 22; Length 419;

Best Local Similarity 70.1%; Pred. No. 1.7e-113;

Matches 295; Conservative 30; Mismatches 85; Indels 11; Gaps 4;

QY 1 MAEPLRGGRPRSRG---GRGARRARGRCPRARQSPARLIPDTVLVLDVSDSDEEVLE 57

DB 1 MAEPVKGKRWGSGGAGRGGRGGRGRRPRAQSPSRGTLDVVSDVLTDSDEILE 60

QY 58 V-----ADPVEVPVARLPAPAKPEQSDSDSEGAAGGAGAPRTLVRRRRRLDPGEA 111

DB 61 VATARGADEVVEPEPPGPGPVASRDNSDSEGEDRRPAGPPREPVRRLVLDPGEA 120

QY 112 PVVPVYSGKVOSSLNLIIPDNSSLLKCPSEPEDEADLTNSGSSPSEDALPSGSPWRKKL 171

DB 121 PLVPVYSGVKVKSRLIPDDLKLLKYPGDEEAEALADSSGLYHEGSPSP-GSPWKTKL 179

QY 172 RKCKEKEKMEFFPDQDISPLPQSSRNKSKKHTKALOKLREVNRKLDLRLVQCKEKH 231

DB 180 RTK-DKEEKKTEFLDNLSPSPSPRPTKSRTHTRALKKLVSEVNRKLDLRLVQCKEKH 238

QY 232 QSPALQSTDEVVLVEGVPVLPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 291

DB 239 QGQEQQOQDEVVVLVEGPTLPETPRLFKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 298

QY 292 VSPNRILLFGESELSPATPTPSTLKLGVADIIDCVVLASSSEATETSOELRLVQCKEKH 351

DB 299 VSPSRILLFGETELSPATPTPRTLKLGVADIIDCVVLTSSEATETSSQQLQVRQCKEKH 358

QY 352 QMLEISLSPDSPKVLMSHYEAMGLSGHKLSFFFDGTTKLSGKELPADLGSGDLEIYVW 411

DB 359 QTLVSLSRDSPKLTLMSHYEAMGLSGKLSFFFDGTTKLSGRELPA DLGMEGSDLEIYVW 418

QY 412 G 412

DB 419 G 419

RESULT 6

AAU10374

ID AAU10374 standard; Protein; 303 AA.

XX

AC AAU10374;

```

XX 14-FEB-2002 (first entry)
XX NF-AT interacting protein 45 variant 3, NIP45V3.
XX
XX NF-AT interacting protein 45 variant; NIP45V; immunosuppressive; human;
XX antiallergic; antidiabetic; antigout; dermatological; antiinflammatory;
XX antiparasitic; antibacterial; vasotrophic; autoimmune; allergy; asthma;
XX nuclear factor of activated T cells; chronic inflammatory disease;
XX allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;
XX vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;
XX diarrhetic colitis; polymyositis; lymphofollicular thyroiditis; gout;
XX dermatomyositis; Sjogren's syndrome.
XX Homo sapiens.
XX WO200181574-A2.
XX
XX 01-NOV-2001.
XX
XX 25-APR-2001; 2001WO-EP04635.
XX
XX 25-APR-2000; 2000US-199356P.
XX
XX (FARB ) BAYER AG.
XX
XX Encinas J;
XX
XX WPI; 2002-055355/07.
XX N-PSDB; AAS17537.
XX
XX Novel nuclear factor of activated T cells interacting protein NIP45
XX variants which can be regulated for treating conditions involving
XX inflammatory processes such as allergy, asthma and autoimmune diseases
XX
XX Claim 1; Fig 10; 107pp; English.
XX
XX The invention relates to an isolated nuclear factor of activated T cells
XX (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and
XX polynucleotide (II) encoding (I). (I) and (II) are useful for screening
XX for agents which decrease the activity of NIP45V polypeptide. (I) is also
XX useful for screening for agents which regulate the activity, preferably
XX cyclic AMP formation, mobilisation of intracellular calcium or
XX phosphoinositide metabolism of human NIP45V. (II) is useful for detecting
XX a polynucleotide encoding a NIP45V polypeptide in a biological sample, by
XX hybridising the polynucleotide to a nucleic acid material of a biological
XX sample. Antibody which specifically binds to (I) or (II) is useful for
XX reducing the activity of NIP45V in a cell, and for detecting (I) in
XX a sample. (I) is useful for modulating the activity of NIP45V in a
XX disease, such as autoimmune, allergic, infectious or chronic inflammatory
XX disease or asthma. NIP45V is useful for controlling diseases including
XX allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal
XX catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,
XX diarrhetic colitis, systemic inflammatory response syndrome, polymyositis,
XX lymphofollicular thyroiditis, sepsis, dermatomyositis, mixed connective
XX tissue disease (MCTD), Sjogren's syndrome and gout. (I) is also useful
XX in diagnostic assays for detecting diseases and abnormalities or
XX susceptibility to diseases related to the presence of mutations in
XX nucleic acid sequence which encode (I). (I) or its fragment is useful
XX in raising specific antibodies. The present sequence represents the
XX amino acid sequence of NIP45V as described in the invention.
XX
XX Sequence 303 AA;
XX
XX Query Match 51.4%; Score 1079; DB 23; Length 303;
XX Best Local Similarity 75.4%; Pred. No. 1e-85;
XX Matches 221; Conservative 21; Mismatches 49; Indels 2; Gaps 2;
XX
XX 120 KVGSSILNIPNSSLKLCPESEDEADLTNSGSSPEDDLPSGPKRLKRRKCEKEE 179
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 13 EVKSLRIIPDLSLTKLYPGEDEEALADSSGLYHGSGSP -GSPKTKLRITK-DKEE 70

```

```

Qy 180 KMEFFPDQDISPLPQSSRNKSKRHTALOKLREVNKRLQDLNSCLSPKQHSALQST 239
Db 71 KKTEFFLDLNSPLSPSPSPRTKSRTHTRALKLSEVNNRRLDLSRCLSPKPGQGOQGO 130
Qy 240 DDEVVLVEGCVLPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVDMHMHGVSPNRLTL 299
Db 131 EDEVVLVEGCVLPETPRPLFPKIRCRADLVRLPLRMSEPLQNVDMHMHGVSPNRLTL 190
Qy 300 LFGSELSPTATPSTLKLGVADIIDCVLASSSEATETSOELRURVQGEKRMLEISLS 359
Db 191 LFGTELSPTATPSTLKLGVADIIDCVLTSSEATETSOQLQRLVQGEKRGQTLEVLS 250
Qy 360 PDSPLKVLMSHYEAMGLSGHKLSPFDGTLKSGKEPLADLGLBSGDLIEVWG 412
Db 251 RDSPLKTLMSHYEAMGLSGHKLSPFDGTLKSGKEPLADLGLBSGDLIEVWG 303
XX
XX RESULT 7
XX AAU10372
XX ID AAU10372 standard; Protein; 334 AA.
XX
XX AAU10372;
XX
XX 14-FEB-2002 (first entry)
XX
XX NF-AT interacting protein 45 variant 1, NIP45V1.
XX
XX NF-AT interacting protein 45 variant; NIP45V; immunosuppressive; human;
XX antiallergic; antidiabetic; antigout; dermatological; antiinflammatory;
XX antiparasitic; antibacterial; vasotrophic; autoimmune; allergy; asthma;
XX nuclear factor of activated T cells; chronic inflammatory disease;
XX allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;
XX vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;
XX diarrhetic colitis; polymyositis; lymphofollicular thyroiditis; gout;
XX dermatomyositis; Sjogren's syndrome.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX MISC-difference 334
XX /note= "Encoded by gc"
XX
XX WO200181574-A2.
XX
XX 01-NOV-2001.
XX
XX 25-APR-2001; 2001WO-EP04635.
XX
XX 25-APR-2000; 2000US-199356P.
XX
XX (FARB ) BAYER AG.
XX
XX Encinas J;
XX
XX WPI; 2002-055355/07.
XX N-PSDB; AAS17534, AAS17535.
XX
XX Novel nuclear factor of activated T cells interacting protein NIP45
XX variants which can be regulated for treating conditions involving
XX inflammatory processes such as allergy, asthma and autoimmune diseases
XX
XX Claim 1; Fig 8; 107pp; English.
XX
XX The invention relates to an isolated nuclear factor of activated T cells
XX (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and
XX polynucleotide (II) encoding (I). (I) and (II) are useful for screening
XX for agents which decrease the activity of NIP45V polypeptide. (I) is also
XX useful for screening for agents which regulate the activity, preferably
XX cyclic AMP formation, mobilisation of intracellular calcium or
XX phosphoinositide metabolism of human NIP45V. (II) is useful for detecting
XX a polynucleotide encoding a NIP45V polypeptide in a biological sample, by
XX hybridising the polynucleotide to a nucleic acid material of a biological

```

CC sample. Antibody which specifically binds to (I) or (II) is useful for  
CC reducing the activity of NIP45V in a cell, and for detecting (I) in  
CC a sample. (I) is useful for modulating the activity of NIP45V in a  
CC disease, such as autoimmune, allergic, infectious or chronic inflammatory  
CC disease or asthma. NIP45V is useful for controlling diseases including  
CC allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal  
CC catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,  
CC diabetoid colitis, systemic inflammatory response syndrome, polymyositis,  
CC lymphofollicular thymitis, sepsis, dermatomyositis, mixed connective  
CC tissue disease (MCTD), Sjogren's syndrome and gout. (I) is also useful  
CC in diagnostic assays for detecting diseases and abnormalities or  
CC susceptibility to diseases related to the presence of mutations in  
CC nucleic acid sequence which encode (I). (I) or its fragment is useful  
CC in raising specific antibodies. The present sequence represents the  
CC amino acid sequence of NIP45V1 as described in the invention.

XX Sequence 334 AA;

Query Match 47.0%; Score 987; DB 23; Length 334;

Best Local Similarity 52.7%; Pred. No. 1.3e-77;

Matches 222; Conservative 24; Mismatches 79; Indels 96; Gaps 5;

QY 1 MAEPLRGGRPRSG---GRGARRARGRCPRARQSPARLIPDTVLVLSVSDSEVLE 57

DB 1 MAEPVGRKRWGSGGAGRGGRGGRAPRAQSPSRGTLDVSVDLVTSDEILE 60

QY 58 V-----ADPEVEVPVARLPAPKPEQSDSDSEGAAGPAGAPRTLVRRRRRLDPGEA 111

DB 61 VATARGAADEVEPEPPGPGVAVASRDNSSDSEGEDRRRPPAGPPREPVRRRRRLVDPGEA 120

QY 112 PVVPVYSGVKVQSSNLIPDNLKPCSPPEDEADLTNSGSSPSDDALPSGSPWRKKL 171

DB 121 PLVPVYSGVKVKSRLIPDNLKLYPPGDEEAEALDSSGLYHEGSPSP-GSPWKTKL 179

QY 172 RKCEKEEKMEFFPDODISPLPQPSNRNKRKHTALOKLRVKNRLQDLRSLSPKQH 231

DB 180 RTK-DKEEKKKTEFLDNLNPSLPPSPRTKSRTHTRALKKLVSEVKNRLQDLRSLSPKPP 238

QY 232 QSPALQSTDDDEVVLVEGVPVLPQSSRLFTKIRCRADLVRLPVMSEPLQNVVDHMANHLG 291

DB 239 QGQEQQEQDEVLVEGTPFTPLPRFPLKIRCRADLVRLPVM----- 282

QY 292 VSPNRILLFGSELSPTATPSTLKLGVADIIDCVVLASSSEATETSQELRLRVOGKEH 351

DB 283 ----- 282

QY 352 QMLEISLSPSPVKVLMHYEAMGLSGHKLSPFFDGTGLSGKELPADLGESGDLIEVW 411

DB 283 -----DSPLKTLMSHYEAMGLSGRKLSPFFDGTGLSGRELPA DLQWESGDLIEVW 333

QY 412 G 412

DB 334 G 334

RESULT 8

AAU10373

XX AAU10373 standard; Protein; 286 AA.

AC AAU10373;

XX 14-FEB-2002 (first entry)

DE NF-AT interacting protein 45 variant 2, NIP45V2.

XX NF-AT interacting protein 45 variant; NIP45V; immunosuppressive; human;

KW anti-allergic; antiasthmatic; antigout; dermatological; antiinflammatory;

KW antipsoriatic; antibacterial; vasotropic; autoimmune; allergy; asthma;

KW nuclear factor of activated T cells; chronic inflammatory disease;

KW allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;

KW vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;

KW diabetoid colitis; polymyositis; lymphofollicular thymitis; gout;

KW dermatomyositis; Sjogren's syndrome.

XX Homo sapiens.

XX WO2000181574-A2.

XX 01-NOV-2001.

XX 25-APR-2001; 2001WO-EP04635.

XX 25-APR-2000; 2000US-199356P.

XX (PARB ) BAYER AG.

XX Encinas J;

XX WPI; 2002-055355/07.

XX N-PSDB; AAS17536.

XX Novel nuclear factor of activated T cells interacting protein NIP45  
PT variants which can be regulated for treating conditions involving  
PT inflammatory processes such as allergy, asthma and autoimmune diseases  
PT -

XX Claim 1; Fig 9; 107pp; English.

XX The invention relates to an isolated nuclear factor of activated T cells  
CC (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and  
CC polynucleotide (II) encoding (I). (I) and (II) are useful for screening  
CC for agents which decrease the activity of NIP45V polypeptide. (I) is also  
CC useful for screening for agents which regulate the activity, preferably  
CC cyclic AMP formation, mobilisation of intracellular calcium or  
CC phosphoinositide metabolism of human NIP45V. (II) is useful for detecting  
CC a polynucleotide encoding a NIP45V polypeptide in a biological sample, by  
CC hybridising the polynucleotide to a nucleic acid material of a biological  
CC sample. Antibody which specifically binds to (I) or (II) is useful for  
CC reducing the activity of NIP45V in a cell, and for detecting (I) in  
CC a sample. (I) is useful for modulating the activity of NIP45V in a  
CC disease, such as autoimmune, allergic, infectious or chronic inflammatory  
CC disease or asthma. NIP45V is useful for controlling diseases including  
CC allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal  
CC catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,  
CC diabetoid colitis, systemic inflammatory response syndrome, polymyositis,  
CC lymphofollicular thymitis, sepsis, dermatomyositis, mixed connective  
CC tissue disease (MCTD), Sjogren's syndrome and gout. (I) is also useful  
CC in diagnostic assays for detecting diseases and abnormalities or  
CC susceptibility to diseases related to the presence of mutations in  
CC nucleic acid sequence which encode (I). (I) or its fragment is useful  
CC in raising specific antibodies. The present sequence represents the  
CC amino acid sequence of NIP45V2 as described in the invention.

XX Sequence 286 AA;

Query Match 37.4%; Score 785.5; DB 23; Length 286;

Best Local Similarity 61.1%; Pred. No. 3.9e-77;

Matches 174; Conservative 23; Mismatches 60; Indels 11; Gaps 4;

QY 1 MAEPLRGGRPRSG---GRGARRARGRCPRARQSPARLIPDTVLVLSVSDSEVLE 57

DB 1 MAEPVGRKRWGSGGAGRGGRGGRAPRAQSPSRGTLDVSVDLVTSDEILE 60

QY 58 V-----ADPEVEVPVARLPAPKPEQSDSDSEGAAGPAGAPRTLVRRRRRLDPGEA 111

DB 61 VATARGAADEVEPEPPGPGVAVASRDNSSDSEGEDRRRPPAGPPREPVRRRRRLVDPGEA 120

QY 112 PVVPVYSGVKVQSSNLIPDNLKPCSPPEDEADLTNSGSSPSDDALPSGSPWRKKL 171

DB 121 PLVPVYSGVKVKSRLIPDNLKLYPPGDEEAEALDSSGLYHEGSPSP-GSPWKTKL 179

QY 172 RKCEKEEKMEFFPDODISPLPQPSNRNKRKHTALOKLRVKNRLQDLRSLSPKQH 231

DB 180 RTK-DKEEKKKTEFLDNLNPSLPPSPRTKSRTHTRALKKLVSEVKNRLQDLRSLSPKPP 238

QY 232 QSPALQSTDDDEVVLVEGVPVLPQSSRLFTKIRCRADLVRLPVRMS 276











```
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52102.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3969-3972; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1279 AA;
XX
XX Query Match 7.4%; Score 154.5; DB 22; Length 1279;
XX Best Local Similarity 23.5%; Pred. No. 0.00042;
XX Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;
QY 6 RRGGRSRGGRGARRGARGRCPRARQSPARLIPDTVLVDLSDEEVLVADPVEVP 65
DB 515 RRGGRGQGFSSPQRASG-----PNRHQAPSMLSPGA---LSSDSKKEGEDEGTEELP 566
QY 66 -----VARLPAPAKPQSDSDSEGAEE-----GPAGAPRTLVRRRRRLL 106
DB 567 ALPVLAKSTKKALASVPSPALPRSLSHWMSRAQESVGLDPAPAAPNPGP---RRGRWV 623
QY 107 DPGEAPVVPYSGVQSLLNLPDSSLLKLCBS--EP-EDEADLTNSGSSPSEDDALPS 163
DB 624 QPG-----VELSVRSMIDLRLQLETLAPSLQDPSQDSLAIIPSGPRKHGQEALET 672
QY 164 G-----SPWRKKLRKK-----EKEE-----KKMEFFPQDIDISPLPQPS---SRNK 201
DB 673 SLTSONEKPRPQASQCSVPHIRLLSQEGVFAQDLEPAPIEDGIVPEPDNPTMDT 732
QY 202 SRXKTEALQK-----LREVNKLQD-----LRCLSPKHQSHPALQSTDDEVIL 245
DB 733 SEFQVQAPARTGLGRVYPGSRSEKHSPPDSACSVDYSSCLSPSEH--PTDSESTEPLS 790
QY 246 VEGPVLQSSRLFTLKRCDRLVLRPMSEPLQNVVDHMANHLGVSPNRILLGGESE 305
DB 791 VDG-----TSSDLEEPAEGBEGGEGGMP-----YGLQE 822
QY 306 LSPTATP-----STLKLGVADIIDCVWLASSEATETSQLRLRVQGEKHKQMLE 355
DB 823 GSP-OTPDQOFLKQHPETLASGAPAP-VQVPERSESRSSRFLQLQVTRPLREP-- 878
QY 356 ISLSPDPLKVLMSHYBEAMGLSGHK 381
DB 879 ---SPSSSSSLALMSRPAPQVPOASGEQ 901
RESULT 15
AAM79953
```

```
ID AAM79953 standard; Protein; 1301 AA.
XX
XX AC AAM79953;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human protein SEQ ID NO 3599.
XX
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX OS Homo sapiens.
XX
XX PN WO200157190-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 05-FEB-2001; 2001WO-US04098.
XX
XX PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK53086.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 393-394; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1301 AA;
XX
XX Query Match 7.4%; Score 154.5; DB 22; Length 1301;
XX Best Local Similarity 23.5%; Pred. No. 0.00043;
XX Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;
QY 6 RRGGRSRGGRGARRGARGRCPRARQSPARLIPDTVLVDLSDEEVLVADPVEVP 65
DB 537 RRGGRGQGFSSPQRASG-----PNRHQAPSMLSPGA---LSSDSKKEGEDEGTEELP 588
QY 66 -----VARLPAPAKPQSDSDSEGAEE-----GPAGAPRTLVRRRRRLL 106
DB 589 ALPVLAKSTKKALASVPSPALPRSLSHWMSRAQESVGLDPAPAAPNPGP---RRGRWV 645
```

```

Oy 107 DGEAFVVPVYSGKVOSSLNLIPNSSLKLCPS--RP-EDADLTNMGSSPSSEDALPS 163
Db 646 QPG-----VELSVRSMDLRQLETLAPLLOPPQSOIALIPSGRKHQGEALET 694
Oy 164 G-----SPMRKTLRKCC-----EKEE---KMEEPDDDISPLPOPS---SRNK 201
Db 655 SLTSONEKPRPRQASQPSYPIHILRLISOEGVAFODLEPAPIEDGIVYEPESDNPMTDT 754
Oy 202 SRKHTALOK-----LREVNKRLOD-----LRCSLSPKXOHPALOSTDDEVVL 245
Db 755 SEFQVQAFARGLGRVYRGSGSSSEHSDSACSVDYSSCLSPRH--PLEDESESTPLS 812
Oy 246 VEGPVLPOSSRLFTLKIRCRADVLVLRPMSEPLQNVVDHMANHLGVSPNRILLFGESE 305
Db 813 VDG-----ISDLDEBPAGDEEBEHEEGGNC-----YGLQE 844
Oy 306 LSPHTP-----STLKGVADIIDCVLLASSBATESTOELRLRVQGEKQOMLE 355
Db 845 GSP-OTPOEOFLKONHFTLLSGAAGAP-VQPERESERSISSRFLLOVOTRPLREP-- 900
Oy 356 ISLSPDSPLYLMSHYEAMGLSGHK 381
Db 901 ---SPSSSLMSRPQAPQASGQ 923

```

Search completed: January 12, 2003, 10:26:29  
Job time : 42 secs

This Page Blank (copy)